

Climate of compromise

The chances of a strong treaty emerging from the United Nations climate talks in Copenhagen seem small, but recent progress offers hope.

With about six weeks left before nations gather in Copenhagen to finish negotiating a climate treaty, hopes are rapidly dwindling that countries will be ready to sign a strong, ratifiable agreement. The pessimism has spread so widely that it could be considered a global pandemic. News stories are already talking about the 'failure' of Copenhagen and squandered opportunities.

But viewed from the perspective of just a few years ago, the Copenhagen summit could already be considered a partial success. In a short span, many nations have pledged to cut their emissions of greenhouse gases by considerable amounts, well beyond any commitments they had made before, such as through the 1997 Kyoto Protocol. Norway, for example, offered this month to reduce its own emissions by 40% below 1990 levels by 2020. Indonesia said it would curb its emissions over that same time by 26% below the levels expected under a business-as-usual scenario, with even stronger cuts possible under an international agreement. The European Union has committed to a 20% reduction below 1990 levels and would increase that to 30% with a global pact. And, for the first time, the US Congress is moving towards establishing laws that mandate emissions cuts.

These words are not to be confused with achievements, but they at least show that countries have started to analyse their own emissions seriously and to develop domestic agendas that would set them on course to meet their commitments. Such unilateral decisions are an essential starting point for an international agreement, and they suggest that countries are now ready to back up their rhetoric in a way that was not true 12 years ago, when they signed the Kyoto Protocol. This is real progress, and it would not have happened without the pressure to produce a treaty.

Nevertheless, such vows fall short of what is needed to protect against the dangers of global warming. Nations need to reduce global emissions far more in

the longer term, and the endgame gets much tougher if leaders delay making those reductions.

In a package of articles this week (see below), *Nature* looks at some of the issues that will play crucial parts in the negotiations in Copenhagen. Several articles focus on factors concerning the developing world, which will endure some of the severest effects of climate change and which will also be responsible for much of the future growth in greenhouse-gas emissions. At the moment, major gaps

remain between the world's wealthiest nations and those still in the process of providing their citizens with basics such as clean water and electricity.

The negotiating impasse can be breached only by concessions on both sides. Developed nations, particularly the United States, must agree to substantial reductions in greenhouse-gas emissions, both in the next decade and in the long term. And developing nations must commit to controlling their greenhouse-gas pollution in some fashion. China has recently taken over as the leader in carbon dioxide emissions and there can be no hope of containing global temperatures without Chinese action.

At the same time, developing nations will need monetary and technical assistance in steering their economies towards a low-carbon future. The wealthy nations have so far committed too little on this front, and the effects of the global recession have tightened budgets around the world. But as economies improve, the wealthiest nations should fashion innovative ways to assist the developing world, whether through the proceeds of carbon trading or through new technical collaborations.

Another major financial obstacle is the issue of support for adaptation. Some estimates suggest that the developing world will require in excess of US\$100 billion in aid every year to cope with the effects of global warming. But the international funds created to help adaptation efforts in the world's poorest nations contain orders of magnitude

EDITORIAL

1027 **Climate of compromise**

NEWS

1034 **Time running out for climate talks**
Jeff Tollefson

NEWS FEATURES

1042 **When the ice melts**
Anjali Nayar

1048 **Counting carbon in the Amazon**
Jeff Tollefson

OPINION

1054 **India pushes for common responsibility**
Rajendra K. Pachauri

1055 **China expects leadership from rich nations**
Jiahua Pan

1056 **Copenhagen needs a strong lead negotiator**
Raúl Estrada-Oyuela

BOOKS & ARTS

1058 **Conveying the campaign message**
Sanjay Khanna

1059 **Q&A: The science of persuasion**
Robert Gifford talks to Sanjay Khanna

1060 **Climate books with clout**
David Reay



For podcast and more online extras see www.nature.com/roadtocopenhagen

less money, and even the available funds have not flowed smoothly to countries in need. The process of distributing funds should be streamlined. But there must be safeguards to ensure that adaptation money is used effectively.

With such major issues still unresolved, pessimistic observers see no chance of success in Copenhagen. But there is still time left for leaders to reach significant agreements if they make it a personal priority and recognize the urgency of the problem. Some leaders, such as British Prime Minister Gordon Brown, have indicated that they would be willing to attend the conference to seal a deal, but more should step forward and they should commit to going. This would lend stature to the negotiations and would raise the chances of achieving a substantial agreement.

It will not be possible to resolve many of the important issues in the remaining time this year. But leaders could make strong progress by building on the momentum at the national level. Many of the commitments made by nations this year are conditional — they depend on other parties taking specific actions as well. These could provide a model for approaching strong targets through a stepwise process.

In the end, successful international negotiations share some important characteristics with scientific research. Both are iterative processes, in which results from one step help to determine the path forward. They require time and perseverance. And they rarely travel in a straight line. Countries should endeavour to build on the positive actions of the past year, both before and after the Copenhagen summit. ■

Russia's grand plan

The creation of a second pillar of excellence will give the country a chance to regain its scientific stature.

When former Russian President Vladimir Putin identified nanotechnology as the spearhead for Russia's economic and technological renewal, critics scoffed, saying that the move was ill-prepared and bound to fail. The multi-billion-dollar initiative would drown in corruption and nepotism, they said. Jokes made the rounds about 'banana' business.

The nanotechnology initiative includes nanoscale techniques with applications in a broad variety of fields, from materials research to mechanical engineering, and from electronics to medicine and biotechnology. Two years on, criticisms of the grandiose project persist. But there is no doubting the political commitment to gingering up Russia's capacities in science, in an attempt to lessen the former superpower's precarious dependence on natural resources.

The creation of the Russian Corporation of Nanotechnologies (Rusnano), a state-owned venture-capital organization that is to shoulder the task of turning Russia into a powerhouse for nanotechnology, is arguably the best thing that has happened to Russian science since the collapse of the Soviet Union. Its procedures for selecting ideas and projects worth investing in are, in terms of thoroughness and transparency, streets ahead of anything else in Russia's dated science-funding system. And in Anatoly Chubais, the 'super-manager' who was responsible for the privatization of Russia's oil and gas industry during the Boris Yeltsin era, Rusnano has a formidable figurehead with the backbone required to negotiate between Russian politics and big business.

No one could complain of a lack of funds: by 2015, Russia will have allocated 318 billion roubles (US\$11 billion) to the initiative (see page 1036). Yet Russia's aspirations could prove over-ambitious because of other critical weaknesses.

In particular, the nation's business and legal environment needs to be more receptive to foreign capital, personnel and know-how. Just as important will be the creation of a research community that can generate ideas and skills: Russia's move into nanotechnology creates a demand for at least 100,000 scientists, requiring in turn a boost in the delivery of Russian undergraduates in relevant disciplines.

The Kurchatov Institute in Moscow, founded during the Second World War to develop nuclear weapons, has been chosen to lead this effort. Diversifying in recent years, it has basic-research competence in most 'nano'-related disciplines. Its director, Mikhail Kovalchuk, has been instrumental in getting the nanotechnology initiative off the ground. But although recently refitted with 6 billion roubles' worth of state-of-the-art equipment for nano-engineering, the Kurchatov Institute cannot master the task alone.

To put basic research in nanotechnologies on a more solid footing, the Russian government is now set to create a national research centre for nuclear physics and nanotechnology. Coordinated by the Kurchatov Institute, the centre will comprise a number of other existing research institutes, such as the Russian Academy of Sciences' Petersburg Institute of Nuclear Physics near St Petersburg.

In addition, the government plans to set up a small number of national research universities. The move, announced earlier this month, is an attempt to overcome the counter-productive separation — a relic of the Soviet science system — of teaching at universities and basic research done mostly at academy institutes.

This move deserves every support.

It has the potential to breathe new life into a research system that has in the past 20 years been plagued by brain drains and frustration, and delivered few internationally recognized achievements. Such a second pillar of basic science, financially and structurally independent from the Russian Academy of Sciences, would be a strong signal of hope, not least to Russia's large scientific diaspora.

But in establishing the centre, the modus operandi of which is still under discussion, the science ministry must avoid the autocracy and subordinating merit to celebrity that hamper many Russian institutions. An international advisory committee should oversee all aspects of the centre's science, recruitment and procurement. Leading positions must be filled with top scientists and administrators, and recruitment decisions should be transparent. Collaboration with leading research institutes abroad, and with Russian university research departments, should be a high priority from the start. Only through such integrity and openness will Russia's drive to science-based prosperity have a hope of succeeding. ■

"The nation's business and legal environment needs to be more receptive to foreign capital, personnel and know-how."

RESEARCH HIGHLIGHTS

BIONANOSENSING

Picky chips

Angew. Chemie Int. Edn doi:10.1002/anie.200902577 (2009)

MicroRNA molecules — key regulators of many cellular and disease processes — can be quickly measured even at low levels using a chip-based detector developed by Shana Kelley of the University of Toronto, Canada, and her colleagues.

The silicon microchip has tiny palladium electrodes coated with synthetic peptide nucleic acids, which bind to specific microRNA (miRNA) molecules. To generate a signal, the researchers incorporated a ruthenium-based reporting system into the chip. When a certain level of miRNA is bound at the electrodes, ruthenium ions are electrochemically reduced, giving off an electrical signal.

The chip was able to detect 100 molecules per microlitre of a specific miRNA in about 30 minutes. It also indicated much higher levels of cancer-associated miRNAs in human head and neck cancer cell extracts than in normal cell extracts.

ATMOSPHERIC SCIENCE

Fire down under

Geophys. Res. Lett. doi:10.1029/2009GL039902 (2009)

The cause of the bush fires that swept across southeastern Australia in February, killing more than 170 people, can be traced back to shifting ocean patterns.

Researchers know that when sea-surface temperatures are cooler than normal in the eastern Indian Ocean, and warmer than usual in the western Indian Ocean, this can result in severe drought in southeastern Australia. Now Wenju Cai of CSIRO in Aspendale, Australia, and his colleagues have linked this ocean anomaly to a heightened risk of fire.

Using data on soil moisture and weather patterns, the researchers found that the Indian Ocean shift preceded 11 of the 21 significant bush fire seasons in Australia



AP PHOTO



Unrealistic

Realistic

Real

Dislikeable likeness

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.0910063106 (2009)

As robots and computer-generated avatars grow more like humans in appearance, real humans become more accepting of them, although only to a point. When non-human figures start looking

too much like humans, people react negatively — an effect called the 'uncanny valley'.

Now Shawn Steckenfinger and Asif Ghazanfar of Princeton University in New Jersey have shown that the uncanny valley is not unique to humans. They presented five macaques (*Macaca fascicularis*) with video footage and static images of real monkey faces along

with realistic and unrealistic synthetic faces (pictured). The macaques preferentially looked at the real faces and unrealistic synthetic ones, shunning the realistic synthetic ones.

As the effect is apparently not dependent on human culture or brain structures, an evolutionary root for this is plausible, the researchers say.

S. STECKENFINGER

since 1950 — including February's fires. These ocean changes have been increasing in number — as might be expected owing to global warming — leading the team to suggest that severe fires could occur more frequently if temperatures continue to rise.

BIOCHEMISTRY

Hungry circadian clock

Science **326**, 437–440 (2009)

Natural rhythms in mammalian physiology and behaviour are set in the brain by the daily light–dark cycle. But in some organs, the circadian clock can be set by nutrient availability — possibly allowing optimal timing of metabolic processes.

Ronald Evans at the Salk Institute for Biological Studies in La Jolla, California, Craig Thompson at the University of Pennsylvania in Philadelphia and their colleagues have identified how nutrient levels influence these light-independent clocks. They found that, in mouse livers, expression of one component of the enzyme AMPK, a nutrient-sensitive metabolic regulator, is cyclical. In addition, AMPK initiates the destruction of cryptochrome 1, a key part of the circadian clock.

Mice in which AMPK signalling was disrupted had altered circadian rhythms.

PHYSICS

Stars in a pellet

Nature Phys. doi:10.1038/nphys1402 (2009)

A team of researchers has used a high-power laser, a tiny hollow plastic pellet and silicon foil to reproduce the X-rays emitted by stars. Shinsuke Fujioka of Osaka University in Japan and his colleagues fired 12 green laser beams at the pellet, compressing and heating it. As the plastic became superheated, silicon atoms near the pellet were ionized, releasing X-rays resembling those seen in two distant binary star systems, where one star is orbiting another.

The authors believe the study could help to refine models for such astronomical systems, and that the radiation from the superheated pellets might recreate the conditions thought to exist at the edge of a black hole.

CELL BIOLOGY

Golgi stretch

Cell **139**, 337–351 (2009)

The Golgi body — a key organelle that processes proteins and lipids for secretion from the cell — is a pile of flattened membranes. Researchers have identified two proteins that work together to maintain the Golgi's shape and function.

Seth Field of the University of California, San Diego, and his colleagues screened proteins to find ones that bind to a particular phosphoinositide that resides in the Golgi membrane and is unique to the organelle. They turned up GOLPH3 and, through a search of its partners, MYO18A. One end of GOLPH3 binds to the Golgi, whereas the other binds to MYO18A, which, in turn, binds to the cell's cytoskeleton. This complex generates tension that keeps the organelle functioning properly and its stretched shape intact.

CANCER BIOLOGY

Metastatic mayhem

Nature Cell Biol. doi:10.1038/ncb1973 (2009)

Motility is crucial for cancer cells to spread and seed new tumours. To learn more about what drives this movement, Erik Sahai of Cancer Research UK's London Research Institute and his colleagues used multiphoton confocal microscopy to follow metastatic breast-cancer cells in live mice.

They found that the cells could travel in groups or individually. Blocking a growth factor called TGF β meant that cells could only move in a group and by means of the lymphatic system. Cells moving alone exhibited activated TGF β signalling and were able to enter the bloodstream. However, TGF β signalling in these cells was transient — although required for single-cell motility, it needed to be shut down to enable the tumour to grow in its new home.

IMMUNOLOGY

Gut response

Immunity 31, 677–689 (2009);

Cell doi:10.1016/j.cell.2009.09.033 (2009)

Hundreds of beneficial bacterial species live in the human gut and help balance the immune system, but which species provide

the benefit and how are not well known. Two research groups have pinpointed only a few select species that seem to do much of the work in boosting immune responses.

The groups measured levels of intestinal T cells in mice after challenging them with various microbes. They found that the most potent immune stimulators were segmented filamentous bacteria. Valérie Gaboriau-Routhiau of INSERM in Paris and her colleagues found that, on their own, these bacteria triggered the same broad T-cell response normally caused by the entire mouse intestinal flora.

Dan Littman of New York University School of Medicine, Kenya Honda at Osaka University in Japan and their colleagues showed that the bacteria induced accumulation of helper T cells, which produce cytokines that mediate a variety of immune responses.

ECOLOGY

Digging diversity

Proc. Natl Acad. Sci. USA doi:10.1073/

pnas.092685106 (2009)

Ant colonies can build large nests with a range of architectures. But what contributes to this diversity? Etienne Toffin and his colleagues at the Free University of Brussels photographed groups of *Lasius niger* ants every 10 minutes for 90 hours as the insects built a nest in a thin layer of sand sandwiched between two glass plates.

The pictures revealed that nest building occurs in two stages. At first, the nest grows quickly in a circular shape. Later, it begins to branch, with buds appearing along the nest wall, some of which go on to become additional chambers. Ant group size affected nest shape; 40% of the 50-ant groups never transitioned to the second stage during the experiment, whereas only 12% of the 300-ant groups remained in the first stage.

PALAEONTOLOGY

Transition fossil

Proc. R. Soc. B doi:10.1098/rspb.2009.1603 (2009)

In the early Jurassic period, about 200 million years ago, the most common flying vertebrates were primitive long-tailed pterosaurs. One hundred million years or so later, the more advanced, short-tailed pterodactyloids came to dominate. How the pterodactyloids evolved from the earlier pterosaurs has been a mystery, but the discovery of a fossil (pictured below)



in China by Junchang Lü of the Chinese Academy of Geological Sciences in Beijing and his collaborators goes some way towards filling in the blanks.

Darwinopterus modularis has the long skull and neck characteristic of pterodactyloids but the rest of the body, including the long tail, is near-identical to that of more primitive forms. The fossil shows the key steps in the evolutionary transition between the two types of organism. It also demonstrates how natural selection acted on groups or 'modules' of characteristics, such as the head and neck, rather than on individual traits.

JOURNAL CLUB

Corinne Le Quéré
University of East Anglia, UK

An oceanographer marvels at the good timing of shrimp.

For many marine organisms, the timing of egg hatching is key to species survival because the time window in which larvae can survive is very short. If eggs hatch too early, they starve before their food source — the spring phytoplankton — blooms. If they hatch too late, they also miss the bloom.

I'm amazed by how often nature gets things right. In most of the North Atlantic, shrimp eggs hatch just a few days before the spring bloom. Peter Koeller of the Bedford Institute of Oceanography in Dartmouth, Nova Scotia, showed that the development and hatching time of shrimp are influenced by local deep-ocean temperature (P. Koeller *et al. Science* 324, 791–793, 2009). This is not surprising, because eggs develop in the deep ocean and their growth rate depends on temperature.

What is surprising is that the shrimp spawn on the right day of the year across the North Atlantic, even though temperatures in the deep ocean vary from one area to the next and do not influence the timing of the spring bloom. Through evolution, the shrimp have adapted to local temperature patterns to spawn at just the right time.

However, this could prove to be a problem for shrimp and the many other zooplankton, fish and shellfish species that have adapted their spawning habits

to local conditions. What will the survival rate of larvae be if deep-ocean temperatures rise, or if the spring bloom occurs earlier? How much time do organisms need to sense and adapt to such changes? These new data will help us to understand the complex interdependence of marine ecosystems, and possibly help to detect potential mismatches between egg hatching and food-source availability.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

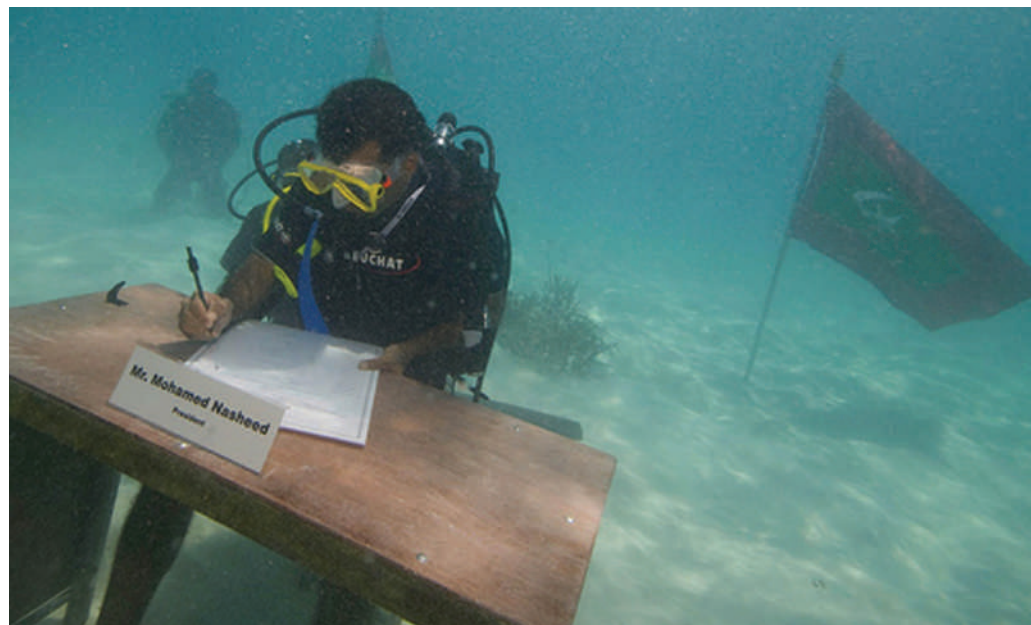
NEWS BRIEFING

● POLICY

European research: Large research and technology networks funded by the European Union (EU) have failed to meet their goal of stimulating additional investment from private businesses, a report published on 14 October by the European Court of Auditors concluded. In evaluating the €17.5-billion (US\$26-billion) Sixth Framework Programme, which ran from 2002 to 2006, the auditors said that such projects generally disintegrate as soon as EU funding runs out and are poorly managed and monitored. Similar criticism was highlighted in March by an independent expert review that called for a radical overhaul of EU research funding. See go.nature.com/n1rEyM for more.

Food security: In a 21 October report, Britain's national academy of science, the Royal Society, urged more investment in agricultural research to stave off a potential global food crisis. The nation should invest at least £2 billion (US\$3.3 billion) over 10 years to fund a research programme studying global food crop security, the report said. Genetic modification of crops, strategic plant breeding and new farming management practices would all be needed to increase yields in a sustainable way.

Energy funding: The US Congress has passed a \$33.5-billion energy and water spending bill for fiscal year 2010. The bill includes \$27.1 billion for the Department of Energy, a 1% increase over the previous year. Of that, \$2.2 billion will go to energy-efficiency and renewable-energy projects, a \$314 million increase from the previous year. But only three of eight proposed 'energy-research hubs' — an idea pushed for by energy secretary Steven Chu — are supported, at \$22 million each. The bill gives hydrogen and fuel-cell programmes



PRESIDENT'S OFFICE, MOHAMMED SEENEN/AP

UNDERWATER COUNCIL

Equipped with waterproof markers, whiteboards and a keen sense of theatre, government ministers from the Maldives held a half-hour underwater meeting on 17 October to highlight the threat of rising waters to their nation's islands. "We should come out of [the United Nations climate summit in] Copenhagen with a deal that will ensure that everyone will survive," said President Mohamed Nasheed (pictured) — although his public-relations stunt is likely to have little impact on stalled negotiations. See page 1034 for more.

\$174 million even though the agency gutted the programmes in its budget request.

Aubergine spiked: The Indian government has delayed commercial release of the country's first genetically modified food crop — an insect-resistant aubergine — despite clearance on 14 October from the nation's top biotechnology regulator, the Genetic Engineering Approval Committee. See page 1041 for more.

Grants cancelled: The Japan Society for the Promotion of Science, an administrative and funding body supported by the science ministry, cancelled a research grant programme on 16 October, after a call by the new government for trimmed budget requests. The 'S-level' Grant-in-Aid for Young Scientists programme would have split ¥2.6 billion (US\$30 million) over the next 5 years among some 30

researchers under 42 years old. The government had previously pared down a massive one-off supplementary budget measure (see *Nature*, 461, 854; 2009), but its new cuts are drawing wider criticism, as the Grant-in-Aid programmes are the yearly lifeline of the Japanese science community.

Resignation: Story Landis, director of the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, resigned abruptly on 17 October from the Interagency Autism Coordinating Committee, a pan-government body coordinating autism research. The move came after the Internet newspaper *Age of Autism* published handwritten notes she left behind after a 30 September meeting of the committee. In the notes, Landis questioned whether one parent on the committee was "pushing autism as [a] multisystem disorder to feed into vaccine injury". Landis

SOUND BITES

"Low-mass planets are everywhere, basically"

Stéphane Udry, University of Geneva

A haul of 32 new exoplanets was revealed on 19 October by a team working at the European Southern Observatory's telescope in Chile's Atacama desert — and Udry says at least as many more will be announced next year. See go.nature.com/WmcxJ3 for more.

apologized in her resignation letter for her “unprofessional” behaviour.

AFP/GETTY IMAGES

● BUSINESS

Solar thermal deal: German engineering giant Siemens announced on 15 October that it would take over Israel's Solel Solar Systems, a leading supplier of mirrors and receivers for solar-thermal power plants, for around US\$418 million. The company says it aims to become the world's market leader in the solar-thermal sector. In March, it acquired a 28% stake in Italian solar-thermal company Archimede Solar Energy, based in Massa Martana.

Insulin in-fighting: The branded biologics industry in the United States, Germany, France, Italy, Spain and Britain could lose US\$6.1 billion in insulin sales to generic competitors between now and 2018, according to a report released last week by Decision Resources, a research firm in Waltham, Massachusetts. Insulins are easier to develop and manufacture than other biologics, and the study presumes that the US Congress will soon adapt the law to create an abbreviated regulatory pathway for biosimilar approvals, as already exists in Europe.

Science park: Plans for a multi-million-pound science park in Britain were unveiled last week by the UK government, the Wellcome Trust and drug company GlaxoSmithKline (GSK). The park will be located at GSK's existing site in Stevenage

NEWS MAKER



Stewart Nozette

The space scientist, who worked for NASA and the US Department of Energy, was arrested by the FBI on 19 October on charges of attempted spying for Israel.

See go.nature.com/bChzoi for more.

and could eventually serve as a base for 1,500 scientists working for early-stage biotech companies. Initial funding has been provided by the UK government (£16.7 million [US\$27 million]), the Wellcome Trust (£6 million), GSK (£11 million) and the East

of England Development Agency (£4 million) — who say they expect eventual investment in the campus to approach £170 million.

● RESEARCH

Charities probe: Paris prosecutors have opened a preliminary investigation into 17 charities, including several that work on medical research into conditions such as diabetes, Alzheimer's disease and cancer. The probe — which under the French judicial system is for information-gathering purposes only — will check what proportion of the millions of euros collected by charities was spent on their stated aims. See go.nature.com/43jhmP for more.

Publishing dispute: The editorial board of the *Proceedings of the National Academy of Sciences* (PNAS) decided last week to publish two papers held up in a dispute with academy member Lynn Margulis, a cell biologist at the University of Massachusetts, Amherst. PNAS editor-in-chief Randy Schekman had written to Margulis about “apparent selective communication of reviews” for a controversial paper by Donald Williamson of the University of Liverpool, UK. The paper, published online in August, is now slated for print. A second paper, co-authored by Margulis, that was also held up, will now be published. See go.nature.com/OuZROE for more.

Ice measurement: A six-year programme of ice-surveying plane flights over Antarctica began on 16 October. NASA's

THE WEEK AHEAD

25–30 OCTOBER

St Louis, Missouri, hosts the 9th International Plant and Molecular Biology congress.

► www.ipmb2009.org

26–30 OCTOBER

The Acoustical Society of America meets in San Antonio, Texas.

► <http://asa.aip.org/sanantonio/program.html>

28–30 OCTOBER

Toronto hosts the Canadian Science Policy Conference, billed as a first step towards creating a robust science-policy network in the country.

► <http://sciencepolicy.ca/mission>

Operation ICE Bridge will provide data on ice thickness and melting to cover a partial gap in satellite readings expected between the time ICESat, one of the ice's observers, comes to the end of its life (probably this year) and a replacement launches in 2014 or 2015.

Conflict-of-interest disclosures:

Twelve journals in the International Committee of Medical Journal Editors — including the *New England Journal of Medicine* and the *British Medical Journal* — have announced they are adopting a uniform format for manuscript authors to disclose their financial interests. See go.nature.com/U7NB4K for more.

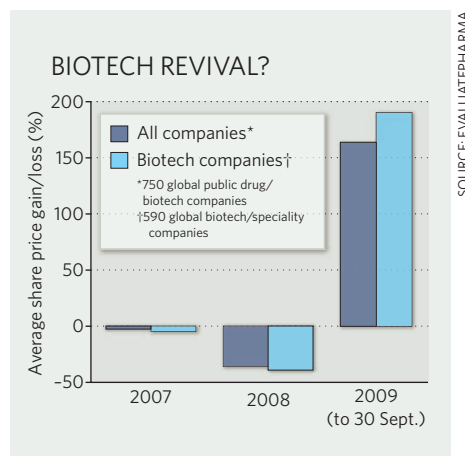
BUSINESS WATCH

In the department of silver linings, a glimmer for the biotechnology sector: the average share price of 360 US biotech companies tracked by EvaluatePharma has risen by 203% since the start of the year. Globally, the year-to-date rise is 163%, in the 590 biotech and speciality companies followed by the London-based research firm. Third-quarter results also show biotech outpacing the pharmaceutical sector in share-price rises (see chart).

The rise is not entirely unexpected. Biotech share prices had almost nowhere to go but up, after risk-averse investors fled the market a year ago at the height of the financial meltdown. “Biotech stocks in particular were

hammered at end of 2008, so you are starting from a really low base,” says Christian Glennie, the senior editorial analyst at EvaluatePharma.

Signs of a true revival may be better gauged by the number of initial public offerings (IPOs). After 41 IPOs in the United States in 2007, there was only one — now delisted — in 2008. This year, there have been three in the United States and another three in the rest of the world. “When you start to see not three biotech IPOs but 33, that will mean that biotech is back,” says Stephen Brozak, president of WBB Securities, an investment research firm in San Diego, California.



NEWS

Time running out for climate talks

Rift between developed and developing nations might be too great.

With just five negotiating days left before the global-warming summit in Copenhagen this December, the international climate community has tempered its expectations and begun to look for a graceful exit.

Pressure on global leaders has been building since negotiators signed the 'Bali Road Map' in December 2007, which set an aggressive timetable that was supposed to conclude in Denmark with a treaty to follow the 1997 Kyoto Protocol. Rich and poor countries alike have spent two years crafting climate policies and proposals, but deep divisions remain.

When irate representatives of developing nations walked out of a meeting with European officials in Bangkok earlier this month, many realized that even technical questions about how to structure a deal might be out of reach this year. Only one interim negotiating meeting remains — in Barcelona, Spain, on 2–6 November. Even staunch optimists are now rethinking their definition of success in Copenhagen.

"I don't like the term 'global deal,'" says Jennifer Morgan, who heads climate policy for the World Resources Institute in Washington DC. "The question is whether we can reach an agreement that allows countries to move forward with confidence."

Many ideas have been floated on how to accomplish that goal. Some argue for building a consensus around areas of agreement and holding off on contentious issues involving money or specific commitments to reduce greenhouse-gas emissions. Others focus on the architecture of a treaty and a broad, if symbolic, agreement among global leaders. Some have suggested simply 'stopping the clock' on negotiations for a few months.

Such proposals are intended to build political confidence and maintain momentum while allowing the United States more time to work out its domestic policy. Although the House of Representatives has passed climate legislation that would curb emissions to 17% below 2005 levels by 2020 and to 83% below by 2050, Senate Democrats are running out of time to pass a companion bill before Copenhagen.

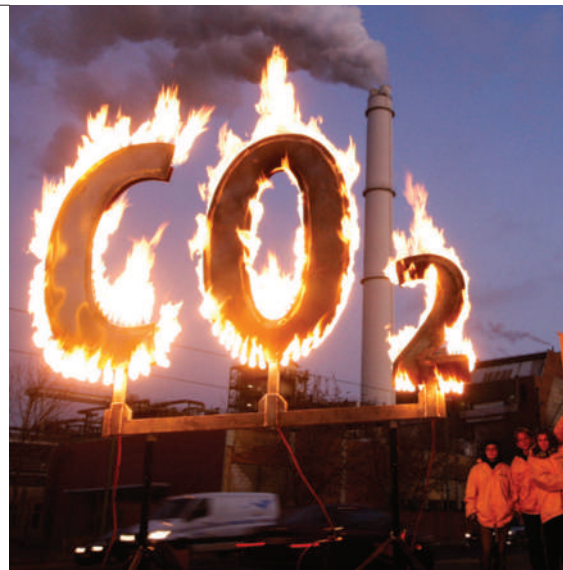
With the United States in limbo, the Europeans are necessarily leading the talks for the

developed world. The European Union (EU) has committed to reduce its greenhouse-gas emissions by 20% below 1990 levels by 2020 and by 30% if there is a global deal, but even those numbers are not enough for most developing countries. The Group of 77 (G77), which represents China, India and the majority of developing nations, is calling on rich countries to cut their emissions by 40% by 2020. So far, only Norway has stepped up to that challenge.

India's environment minister, Jairam Ramesh, says the negotiations are unlikely to go anywhere unless wealthy nations embrace more ambitious emissions reductions and promise more money to help developing countries cope with climate change. He says negotiators should focus on a few key issues, including deforestation, adaptation money for the least developed nations, and the Clean Development Mechanism, which allows rich countries to offset their emissions by paying for clean-energy and other 'green' projects in the developing world.

"We need a more limited and more pragmatic approach if we are going to make Copenhagen a success," Ramesh says.

At the same time, wealthy nations are insisting on stronger commitments from developing countries — many of whom have responded by introducing significant new policies. Indonesia recently announced plans to reduce its emissions by 26% by 2020 compared with predicted levels if emissions were left uncurbed, pledging a 41% reduction if it receives international aid for measures such



as stopping deforestation. Even India, which has repeatedly said it won't commit to anything until its per-capita emissions reach those of the developed world (see graphic), has proposed the idea of 'voluntary mitigation outcomes' that would allow the country to calculate specific emissions reductions based on expanding solar power and other existing climate policies.

China, the world's largest emitter, is moving forwards with aggressive energy-efficiency targets and renewable-energy mandates — but has yet to pledge binding commitments or agree a date to level off its explosive emissions growth.

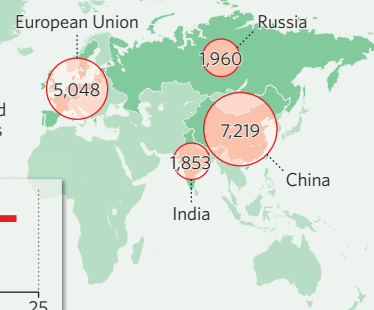
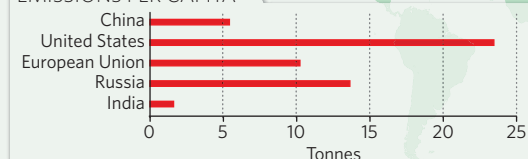
Many were caught off guard in Bangkok when officials from the G77 and Europe fought over the future of the Kyoto Protocol. The G77 wants to maintain Kyoto — which is binding for developed but not developing countries on greenhouse-gas emissions — as the primary international framework. EU officials want to create a single new treaty under the original UN Framework Convention on Climate Change and then import the Kyoto Protocol. The United States claims neutrality, but some fear that it could impede progress by cherry-picking which parts of Kyoto it would



FIVE BIGGEST EMITTERS

Total greenhouse-gas emissions, excluding land-use change, in millions of tonnes of carbon dioxide equivalent for 2005

EMISSIONS PER CAPITA



AP PHOTO/M. SCHREIBER

SOURCE: WORLD RESOURCES INSTITUTE



be willing to subscribe to in a new treaty.

Some observers believe that the biggest challenge will be money, particularly in the midst of an economic crisis. Costs for adaptation to climate change in the developing world alone have been estimated at US\$100 billion annually, and some suggest the figure could be two or three times that amount. The World Bank estimates that establishing a low-carbon economy in developing countries could cost several hundred billion dollars each year.

Money could enable progress on other fronts. Some developing nations have proposed to allow rich countries that cannot sufficiently reduce their emissions to make up the shortfall with cash. Indeed, after blasting the inadequacy of the US climate proposals, India's Ramesh smiles and acknowledges that a deal would be much more likely if the United States were to put, say, \$200 billion on the table.

Saleemul Huq, who tracks adaptation issues for the London-based International Institute for Environment and Development, says that developed countries will have two more chances to address the issue before Copenhagen (see also *Nature* doi:10.1038/news.2009.1019; 2009). The EU finance ministers were scheduled to meet as *Nature* went to press, and a meeting of the G20 finance ministers is due in early November.

"One expects that they will come back with some kind of offer, and that would then unlock other issues in the negotiations," Huq says. "They will still be wrangling about the amounts, but at least they will go into the process of negotiating. Right now they are posturing, not negotiating."

Jeff Tollefson

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen

Hwang verdict imminent

Despite his research being exposed as fraudulent and unethical almost four years ago, the career of South Korean cloner Woo Suk Hwang has thrived. He has established a research institute, laid claim to a set of human-cloning patents, received a scientific excellence award, published a handful of papers and entered into a collaboration with a powerful provincial government.

Now Hwang faces his biggest hurdle yet: the possibility of jail time. As *Nature* went to press, a verdict was expected on charges that he knowingly used faked research results to apply for grants, embezzled as much as 2.8 billion Korean won (US\$2.4 million), and purchased human eggs in violation of the country's bioethics law.

The charges relate to two papers in which Hwang claimed to have created stem-cell lines from cloned human

embryos, and to have used the technology to develop patient-specific stem cells (W. S. Hwang *et al. Science* 303, 1669–1674; 2004; and W. S. Hwang *et al. Science* 308, 1777–1783; 2005). Both were found to have been fabricated.

In August, at the last of 43 judicial hearings on the affair, prosecutors sought four years in prison for Hwang. It was the end of an epic 40-month trial that saw the chief judge change three times, heard testimonies from more than 60 witnesses and included some bizarre moments — such as when Hwang claimed he could not account for some of his funding because he had used it to pay Russian mafia for access to the DNA of a frozen mammoth that he hoped to clone.

Hwang still has powerful supporters. In June, he received an award from the Jang Yeong-sil Memorial Foundation for scientific excellence. In August, the South Korean province of Gyeonggi announced that it planned a collaboration with Hwang to develop transgenic animals able to grow organs for transplantation to humans. And on 12 October, 33 parliamentarians submitted a petition to the Seoul Central District Court Criminal Division pleading for leniency in sentencing Hwang so that he can continue

his research without further ado.

Meanwhile, stem-cell science in South Korea has also bounced back from the scandal. In July, the government promised to triple its funding for the field to 120 billion won by 2015. The government is also putting together a plan to give about 20 top stem-cell projects up to 1 billion won each over two years, with the possibility of twice that amount in a subsequent funding round. In April, the government approved an application by Seoul's Pochon CHA University to carry out human therapeutic cloning experiments.

"The future of stem-cell research is very promising," says Dong-Youn Hwang of the Pochon CHA University. "Many Korean people now realize that stem-cell-based therapy is not magic, but only results from hard work, huge investment and continuous support," he adds.

Developmental biologist Kunsoo Rhee of Seoul National University, where Woo Suk Hwang worked until he was dismissed in 2006, says that the increased attention on scientific conduct has had some positive outcomes, including an undergraduate course on misconduct, which began at the university two years ago.

"If we have something like [the Hwang case] all the time it would be bad, but once every 10 years — or hopefully only once in our lifetime — it can be a learning experience."

"After Hwang's scandal, Korean biotech is getting more transparent," adds Kyung-Sun Kang of Seoul National University's Adult Stem Cell Research Center. Kang says that although research on embryonic stem cells was difficult for a while, there has been an upsurge in the commercialization of adult stem cells, with several stem-cell therapies now in clinical trials.

If Woo Suk Hwang avoids a prison sentence, he says that he plans to continue his efforts to benefit South Korea by continuing his research on mammalian cloning, and still hopes to vindicate his human-cloning technique.

David Cyranoski



UK FACES TOUGH FUNDING QUESTIONS
Universities told international collaborations will be key.
go.nature.com/ACNYtr

CERN/M. BRICE



Hwang is charged with fraud.

CHUNG SUNG-JUN/GETTY

High hopes for Russia's nanotech firms

But an ambitious government initiative has been slow to incubate a domestic high-tech industry.

MOSCOW

Russia's investment in the very small is about to get bigger. The government announced this month that it is injecting 318 billion roubles (US\$11 billion) into its ambitious plan to develop and commercialize nanotechnologies.

Over the past two years, the Russian government has spent nearly 94 billion roubles on the effort. The goal is to establish a self-supporting domestic high-tech industry by 2015, to help reduce the country's dependence on oil and gas exports. But so far the project has created only 2,000 or so new jobs and a small number of public-private ventures.

"There has been little success so far," Russian president Dmitry Medvedev said at the opening of the International Nanotechnology Forum in Moscow on 6 October. "But we do have the knowledge, the financial resources and the administrative capacity to become leaders in a technological process that will change the world."

The financial crisis has sharpened the impetus for Russia to diversify its industries. Oil and gas prices dropped and foreign investment in Russia plunged by 31% between the first half of 2008 and the first half of 2009.

Russia's ambitious nanotechnology initiative became federal law in 2007 after Mikhail Kovalchuk, director of the Kurchatov Institute in Moscow, convinced then-President Vladimir Putin of the economic potential of the burgeoning field (see *Nature* 448, 233; 2007). Business analysts Lux Research expect global revenue from nanotechnology products to increase tenfold, from US\$250 billion to \$2.5 trillion, over the next 5–6 years; applications range from cosmetics and innovative drugs to low-friction coatings and environmental sensors.

Russia is catching up with other world leaders in its public investment in nanotechnology (see graph), but it is still a minor player in making products. At current exchange rates, domestic sales and exports of Russian-made nano-enabled products are estimated at \$4 billion to \$5 billion for 2009 — no more than 2% of the global market volume.

Tasked with changing that is the state-owned venture-capital company, the Russian Corporation of Nanotechnologies (Rusnano).



Exhibits at a major nanotech forum in Moscow this month highlighted the field's economic potential.

By 2015, Rusnano aims to have attracted 450 billion roubles in private capital, and generated sales of 900 billion roubles, chief executive Anatoly Chubais told the Moscow conference. "We are still in the valley-of-death curve" between product development and commercialization, he says. "But our newborn baby is growing nicely."

But business analysts remain sceptical as to whether the corporation will be able to meet its goals. "We're seeing a real and serious commitment to making Russian industry more innovative," says Michael Holman, an analyst with Lux Research in New York.

"I'm impressed by the steps they have taken. But it's nonetheless going to be an uphill struggle, not least because foreign investors are not always comfortable with doing business in Russia."

Russia's nanotechnology work, in its broadest definition, involves 400 or so academic science departments and some 200 companies, says Leonid Gokhberg, an expert on econometrics at the Higher School of Economics in Moscow. The weakness of Russia's science system, which bars most university academics from applying for research funding and doesn't encourage entrepreneurial activity, is a barrier to efficient technology transfer, he says.

Creating decent business conditions and a more competitive academic environment

is the next challenge, Medvedev told the nanotechnology forum. To that end, Rusnano says that it will lessen the initial investment risk for entrepreneurs by signing on as a co-investor, minority shareholder or supplier of cheap loans. In such joint ventures, any profits would be shared between the project applicant, the private investor and Rusnano, but Rusnano will claim no ownership or intellectual-property rights.

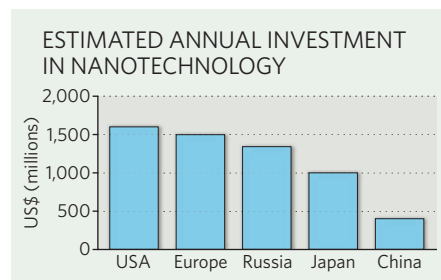
On the right track

Since its creation, Rusnano has received 1,200 project proposals, but most have not met its requirements. "At first, few people understood what we are looking for," says biologist Sergey Kalyuzhnyi, one of Rusnano's board members and director of its department for technical expertise. "In most cases we had to say 'sorry guys, nice idea, but not enough demonstration of market potential!'"

But awareness is increasing, he says. After a two-stage review process, Rusnano has now selected 36 projects for funding, all of which



President Dmitry Medvedev (right) and Rusnano's chief executive, Anatoly Chubais.





FLIES GET FRIGHT FROM FALSE MEMORIES

Light activation pinpoints where learning happens in fruitflies.

go.nature.com/gMnKEv

SPL

involve commercial production at manufacturing facilities in Russia. They include the Novosibirsk-based company Sun Innovations, which produces nano-ink for printers, and Rybinsk-based Saturn, which makes nano-structured coatings for use in aircraft.

Russia is also keen to attract international partners, which has proved difficult. Only two projects involving foreign partners have been approved so far, including a 1.1-billion-ruble joint venture with the German firm VI Systems in Berlin. The company, which develops laser-based components for ultra-high-speed computers, networks and interconnects, will start production in 2011 at a newly built facility in St Petersburg.

"It's a perfect match for us," says company founder Nikolay Ledentsov, a physicist formerly with the Ioffe Physical Technical Institute in St Petersburg. VI Systems currently manufactures its devices in Taiwan and elsewhere. "In the future we can safely produce in Russia, itself an emerging market for electronics, and share profits," he says.

Rusnano is also participating in joint infrastructure projects with firms such as RusChemBio, a Moscow-based 24-hour delivery service that supplies imported compounds and reagents required for clinical diagnostics, chemistry and drug development. Several drug companies, including Roche and Aventis, said at the forum that they will discuss setting up joint ventures in drug development with Rusnano.

Russia's push into nanotech is part of its larger effort to strengthen basic science and education. The Kurchatov Institute, Russia's premier centre for nuclear physics, last year received more than 5.9 billion roubles from the government to set up a new centre for nano-, bio-, info- and cognitive sciences, including a state-of-the-art genome-sequencing facility, an upgraded synchrotron radiation machine and a new data-processing centre. And concrete plans exist for a national research centre for nanotechnology that includes the Kurchatov and several research centres run by the Russian Academy of Sciences.

But experts warn of exaggerated hopes. "Russia is throwing a lot of money and brain power at nanotechnology," says Ian Miles, an expert on technological innovation at the University of Manchester, UK. "But favourable conditions, from more favourable starting points, are being created everywhere. It's not going to be easy."

Quirin Schiermeier

See Editorial, page 1028.

Hazy goals hold up conservation

CAPE TOWN

With the world almost certain to miss a target to slow extinction rates by 2010, governments are looking to adopt new biodiversity targets next year. What those might be — and what science will be needed to underpin them — is yet to be established, researchers said last week at a conference on biodiversity held in South Africa.

"2010 may not be the year when we reverse the rate of loss of biodiversity, but needs to be the year when we reversed the response to that loss," says Achim Steiner, executive director of the United Nations Environment Programme (UNEP).

In 2002, more than 120 countries adopted a target to achieve a "significant reduction" in the rate of biodiversity loss by 2010. That aim will not be met, says Georgina Mace, director of the Centre for Population Biology at Imperial College London. But just how much the world will miss the target by is difficult to calculate, because the target does not specify a baseline extinction rate from which to start counting. "The lack of baselines and timescales are quite problematic," Mace says.

The next generation of targets will aim for a more positive outcome and set more easily measurable goals, says David Cooper of the Convention on Biological Diversity secretariat in Montreal. The targets are likely to aim for a complete halt to biodiversity loss by 2050, and to set more modest interim targets for 2020.

The new targets will be designed to encourage countries to address the underlying drivers of biodiversity loss, such as climate change and unsustainable land use, Cooper says. They will also recognize the socioeconomic value of 'biodiversity services', such as tourism revenue generated by coral reefs or the carbon sequestration value of a forest.

The new goals are being agreed through a set of international negotiations, to culminate in Japan in October 2010. There, governments will also consider whether to set up an Intergovernmental Platform on Biodiversity and Ecosystem Services to bridge the gap between science and policy.

Science advice is currently slow to reach policy-makers, says Steiner, and

too piecemeal when it gets to them. "The number of scientific assessments, their focus and assumptions, are simply too bewildering and fragmented," he told the Cape Town conference.

Scientists also need to find better ways of estimating biodiversity loss, Mace says. Writing in *Science* last month, a team led by UNEP's Matt Walpole identified serious shortcomings in the indicators used to measure progress towards the 2010 target, such as monitoring changes in species' status on the International Union for Conservation of Nature's endangered 'red list' or in the size of protected areas. But the indicators do not include any measurement of the effects of climate change on biodiversity, and few address the societal benefits of safeguarding species (M. Walpole *et al. Science* 325, 1503–1504; 2009).

Cooper adds that scientists need to better integrate data to provide a more holistic



The red list names nearly 700 frogs and toads as 'endangered'.

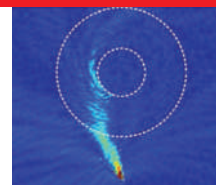
picture of biodiversity trends. "We need to develop a monitoring system," he told *Nature*.

The Biodiversity Observation Network of the Group on Earth Observations offers such a system for 'joining the dots' of global biodiversity knowledge. Launched in 2008, the network aims to be a one-stop shop for information on biodiversity. It will release indicators, generate maps of priority areas, provide conservation plans and report on trends in uses of biological resources.

Chairman Bob Scholes, of South Africa's Council for Scientific and Industrial Research, says the network will start producing a "strong element" of this bigger picture in two to three years' time.

Linda Nordling

DIGITALSTOCK



**RESEARCHERS CREATE
PORTABLE BLACK HOLE**
Metamaterials ensnare
microwave light.
go.nature.com/w1g2Mf

Europe's Galileo project gains ground

Long-troubled satellite-navigation system receives formal backing from European Commission president.

BRUSSELS

Europe is inching closer to realizing its far-reaching plans for Earth-observing. Last week José Manuel Barroso, president of the European Commission (EC), publicly confirmed his support for the Galileo and GMES satellite-navigation and Earth-monitoring programmes.

"Without space applications we will not be able to observe and tackle climate change," Barroso told a European space conference held in Brussels on 15–16 October. "Without space research, our knowledge society will simply not come about."

Critics point out, however, that Galileo has been in the works for a decade and still does not have the first of 30 planned satellites in orbit.

Volker Liebig, director of Earth-observation programmes for the European Space Agency (ESA) in Frascati, Italy, said that Barroso's presence at the meeting was "a real sign of change". It is the first time that Barroso, who was re-elected in September for a second five-year term as president, has attended a meeting on Galileo and offered his support so publicly. "This is proof of the stronger and stronger involvement of the European Union [EU] in space policy," says Alain Bories, director for strategy at satellite manufacturer OHB Technology in Bremen, Germany. "This is good news."

Galileo, Europe's ambitious answer to the US Global Positioning System (GPS), is set to become operational in 2013. Its troubled past includes several near-cancellations of the project, including a 2006 collapse of the public-private partnership intended to fund it. The project was revived in 2008, when it was decided that the EC would fund Galileo completely, using surplus money from the agriculture budget. Two test satellites for the system have been launched, in 2005 and 2008, although the latter was two years late.

Barroso told the conference that Europe can succeed in getting Galileo launched. "We must guarantee the success of our flagship programmes," he said. "Governance issues must not get in the way." Space policy will be further entrenched in EU policy if the Lisbon treaty is ratified as expected by January 2010; the treaty contains a provision stating that a pan-European space policy will be drawn up.

ESA developed Galileo and is contracted to procure services for it. This process is now



Europe plans to launch up to 30 satellites that will be independent of the US Global Positioning System.

in full swing, with a 13 November deadline for final tenders from companies bidding to provide satellites, integration systems and launches. "The plan is to sign the contracts as soon as possible," says Karamitsos.

The two leading contenders to build the bulk of the satellites for Galileo are EADS Astrium and a partnership between UK space company Surrey Satellite Technology, based in Guildford, and Germany's OHB. "There's a danger Europe will miss the boat," says Philip Davies of Surrey Satellite. "We're 10 years into the programme and still don't have working satellites."

The first four satellites are slated for launch by the end of next year, although this phase of the project has already overrun its budget by €376 million (US\$560 million). "In all space programmes there's always an overspend," argues Fotis Karamitsos, who directs the Galileo efforts for the EC's directorate-general for energy and transport. The money must come out of the overall €3.4-billion budget for Galileo and the related navigation system EGNOS (European Geostationary Navigation Overlay System), which became operational on 1 October.

The remaining 26 satellites are slated to begin launching at the end of 2012. But the current launch capacity is two satellites per launch, with a maximum of four launches per year from the Kourou site in French Guiana. That would mean that the maximum number of satellites in space by the end of 2013 would

be only 14. An updated Ariane launcher capable of launching four satellites at a time might be ready by 2013, says René Oosterlinck, ESA's Galileo director.

While politicians discussed funding and launch details in Brussels, scientists gathered in Padua, Italy, last week to talk about what they can learn from Galileo.

Because both the GPS and Galileo emit signals that are reflected by Earth's surface, these signals can monitor long-term changes on the planet. "The biggest interest is that [scientists] have 40 or 50 satellites moving around, guaranteed, for years and years emitting signals," says Oosterlinck. Potential investigations, he says, could range from testing Einstein's theory of relativity to probing the ionosphere and troposphere. Each of the main satellites can also carry a small payload; around half of the satellites might be able to carry a small project for outside researchers, says Oosterlinck.

Meanwhile, Europe's GMES (Global Monitoring for Environment and Security)/Kopernikus project might benefit from lessons learnt from Galileo's funding tribulations. Five families of satellites, known as Sentinels, are to be launched, beginning in 2012. So far, €1.6 billion of the estimated €2.7-billion budget has been raised, including €831 million allocated at an EU ministerial meeting last November.

But without a better way to finance big projects over many years, Liebig says, GMES might fall prey to the same funding turbulence that has plagued Galileo.

Katharine Sanderson

**"We must guarantee
the success of our
flagship programmes."**

Fossil primate challenges Ida's place

Controversial German specimen is related to lemurs, not humans, analysis of an Egyptian find suggests.

A 37-million-year-old fossil primate from Egypt, described today in *Nature*¹, moves a controversial German fossil known as Ida out of the human lineage.

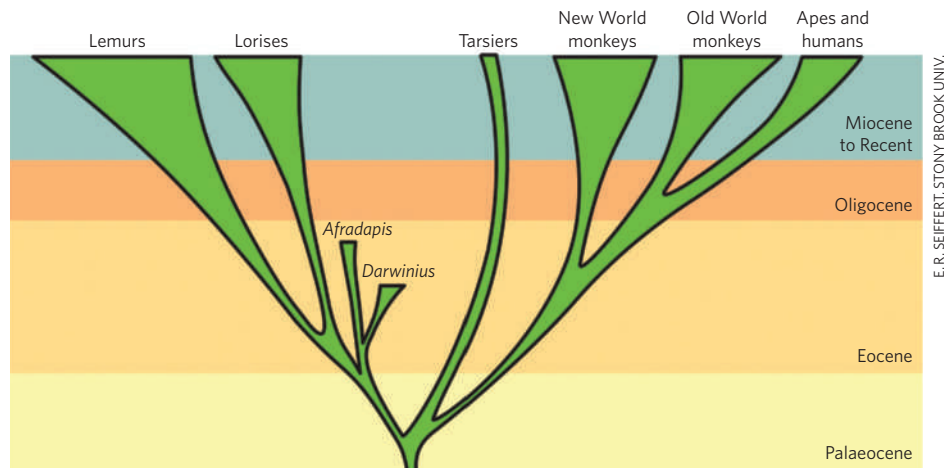
Teeth and ankle bones of the new Egyptian specimen show that the 47-million-year-old Ida, formally called *Darwinius masillae*, is not in the lineage of early apes and monkeys (haplorhines), but instead belongs to ancestors (adapiforms) of today's lemurs and lorises.

"Ida is as far away from the human lineage as you can get and still be considered a primate," says Christopher Beard, a palaeoanthropologist at the Carnegie Museum of Natural History in Pittsburgh, Pennsylvania, who was not involved in either research team.

Philip Gingerich of the University of Michigan in Ann Arbor, a lead author on the Ida report, said in an e-mail that his group continues to stand by its analysis that their specimen is closer to monkeys and apes than to lemurs. Led by Jørn Hurum, a researcher at Oslo's Natural History Museum, the team promoted the near-complete Ida skeleton as a missing link in human evolution, although the formal publication, in *PLoS ONE*², didn't go that far. Adapiforms are not in the human lineage.

After years of secret deals to purchase the skeleton and prepare the fossil, Ida was launched in May in a lavish entertainment campaign³ that included a book and film projects. New York City mayor Michael Bloomberg was enlisted for her unveiling at the American Museum of Natural History.

Ida's skeleton is the earliest such primate specimen found, and offers great insight into the development of these animals long before apes and humans shared a last common ancestor — estimated at 8 million to 10 million years ago. But immediately upon publication, the conclusions about Ida were roundly criticized⁴ by leading early-primate researchers, including Erik Seiffert of Stony Brook University in New



Darwinius was proposed to be on the lineage to humans; instead, it and *Afradapis* are on another branch.

York and Elwyn Simons of Duke University in Durham, North Carolina.

Seiffert and Simons are co-authors of today's *Nature* report on the new species, *Afradapis longicristatus*.

Probing sand-buried sediments about 60 kilometres southwest of Cairo, in 2000 their team found the first specimens of a series of teeth and jaw bones. Using delicate brushes, over the years they uncovered the near-complete dentition, and were close to submitting it for publication when Ida rocketed into the headlines, even garnering its own Google logo.

Seiffert says that, given Ida's wonderful preservation and undisputed age, his team needed to include it in their planned publication analysis, which examined more than 117 living and extinct primate species records. This analysis not only defined the context for *Afradapis*, but put both *Afradapis* and *Darwinius* in the adapiform lineage.

As part of this added analysis, Seiffert travelled to Gingerich's Michigan lab last spring to examine a high-resolution cast of Ida. Seiffert says that characteristics said to link Ida to haplorhines weren't supportable. Key among these was an ankle bone, the talus, which Gingerich's group reported as straight-sided, like that of a haplorhine species. "It is not possible to say this because the specimen is damaged and the bone obscured," says Seiffert.

Gingerich says the differing interpretation may arise because a cast, not the original, was examined. He noted that his group is running computerized tomography scans of Ida to more clearly describe the fossil's feet and hands. In addition, Hurum says, "*Darwinius*

as a skeleton is much more complete than *Afradapis*, and *Darwinius* shows additional haplorhine characteristics not preserved here".

In the *Nature* report, the authors describe dental characteristics that place their species as an adapiform. The last molar is relatively long from front to back, and the front cusp of each molar is missing — both adapiform traits. "We discovered a close relative of Ida," says Seiffert. "But they both are in the lemur lineage."

Contacted at the dig site in the Fayum Depression region late last week, Seiffert said the team had uncovered more ankle bones of their species that confirm their *Nature* diagnosis. The team now has nearly 40 *Afradapis* individuals. The animal was the largest primate of the time, weighing nearly 3 kilograms and thriving on insects and fruit.

But their animal doesn't have a nickname. No popular-science book has been published. And no film project is planned. It seems unlikely that *Afradapis* will get its own Google logo. ■

Rex Dalton

1. Seiffert, E. R. *et al.* *Nature* **461**, 1118–1121 (2009).
2. Franzen, J. L. *et al.* *PLoS ONE* **4**, e5723 (2009).
3. *Nature* **459**, 484 (2009).
4. Simons, E. L. *et al.* *Nature* **460**, 456 (2009).

ANNOUNCEMENT

Alison Abbott, *Nature*'s senior European correspondent, has been awarded the Euroscience Foundation's European senior Science Writers' Award for 2009. Abbott received the €3,000 (US\$4,500) prize "for being continuously committed to high quality science and research policy journalism" at a 14 October ceremony at the University for Music and Drama in Hanover, Germany.



Years of painstaking research revealed the fossil.

Transgenic aubergine put on ice

Stiff opposition from activists has persuaded the Indian government to put off commercial release of the country's first genetically modified (GM) food crop, despite clearance from the nation's top biotechnology regulator.

The 14 October ruling by the Genetic Engineering Approval Committee (GEAC) granted permission for Indian farmers to grow a transgenic version of aubergine, or brinjal, that is insect-resistant. But barely 24 hours later, Jairam Ramesh, India's minister of environment and forests, said that permission for its cultivation will be given only after consulting "all stakeholders".

Ramesh says that the ministry will seek public comments until the end of the year and that he "will have a series of consultations with scientists, agriculture

for Cellular and Molecular Biology in Hyderabad, says Ramesh has made the right choice. "The government need not accept every recommendation made by the GEAC," he says. Bhargava was one of the three members of the GEAC, out of a total of 20-odd members, who opposed the introduction of *Bt* brinjal — citing what they called inadequate safety data provided by Mahyco.

Mahyco says that at least 25 environmental-safety and food-safety studies on animals carried out since 2002 show that *Bt* brinjal is "absolutely safe" to eat. But Bhargava and activist groups argue that the GEAC did not get the company data independently analysed. The only other study, by French scientist Gilles-Eric Seralini of the Committee for Independent Research and Information on Genetic Engineering, branded *Bt* brinjal "potentially unsafe for human consumption".

According to Seralini, eating *Bt* brinjal reduced appetite in goats, increased prothrombin time (the time it takes blood to clot) in goats and rabbits, and caused the plants to produce a protein inducing resistance to the antibiotic kanamycin. However, an expert committee dismissed these concerns, saying that the crop "has been extensively tested for its biosafety, and no additional studies/review are necessary". That expert report formed the basis for the GEAC's ruling.

The Coalition for a GM-Free India called the approval a "shame" and alleged that "regulators have put the interests of corporations over that of ordinary citizens". But Rao says the anti-GM lobby is nervous. "They have already lost the battle over *Bt* cotton — the only GM crop grown in India — and they know if they lose over *Bt* brinjal they lose the war," he says. ■

K. S. Jayaraman



P. PARANJPE/REUTERS/CORBIS

India is mooting commercial use of GM aubergines.

experts, farmers' organizations, consumer groups and NGOs" in January and February 2010 before deciding whether to go forward.

The GM brinjal variety was developed by Mahyco Monsanto Biotech, a joint venture between Jalna-based Maharashtra Hybrid Seed Company and US seed giant Monsanto.

The decision to seek further input has angered some crop scientists. "The minister has set a bad precedent by ignoring the recommendation of the GEAC — a statutory body consisting of scientists," says Chavali Kameswara Rao, secretary of the Foundation for Biotechnology Awareness and Education in Bangalore. "The biosafety issue of *Bt* brinjal has been studied by more than 150 scientists, and nothing new will come from fresh consultations."

But GEAC member Pushpa Bhargava, who was founding director of the Centre

Corrections

The News story 'Where the US stimulus money is going' (*Nature* **461**, 856–857; 2009) gave the wrong location for the National Synchrotron Light Source II. It is at the Brookhaven National Laboratory in New York. And in the News Feature 'The disappearing nutrient' (*Nature* **461**, 716–718; 2009), the estimate for the amount of phosphate that could be extracted from the ground should have been 47 billion tonnes not 62 billion. This means that the figures in the pie chart should have read: United States (3.4 Gt), China (10 Gt), Morocco and Western Sahara (21 Gt), South Africa (2.5 Gt) and Rest of the world (10.1 Gt).



When the ice melts

Deep in the Himalayas, the disappearance of glaciers is threatening the kingdom of Bhutan. **Anjali Nayar** trekked through the mountains to see how the country is adapting to a warming world.

Kaka Tshering loops a piece of frayed jute rope around a 150-kilogram boulder. A handful of his fellow workers line up on either end and pull the rope taught.

"Shochi, Shoni," the workers call in unison, as they heave. Their voices are raspy from the 4,400-metre altitude and moist, cold air. "Put your strength together."

After rocking a couple of times, the boulder rolls over and the labourers tumble backwards. Their cries are drowned out by the furious work around them as more than 300 men and women scrape away rocks with spades and shovels to reach their daily quota.

The work force is a cross-section of life in Bhutan. There are a number of young drop-outs from the capital Thimphu, with greasy, shoulder-length hair and tattoos running up their forearms. Retired soldiers from the Royal Bhutan Army labour alongside former students of Buddhism. There are a handful of women in their traditional tartan-style robes, beaded necklaces and antique silver and turquoise brooches.

Together in matching hard hats and leaky rubber boots, they make up Bhutan's army against the effects of climate change. Their task

is to deepen and widen the outlet channel from lakes formed by the rapidly melting Thorthormi glacier (pictured above, to the right of the lake). By helping the water to drain faster, Bhutanese officials hope to prevent a catastrophic flood.

Glaciers in the Himalayas are retreating faster than in any other part of the world and they could disappear completely by 2035 (ref. 1). This puts the mountainous nation of Bhutan at a special risk. In an area smaller than Switzerland, it has 983 glaciers and 2,794 glacial lakes, some of which have burst to produce deadly glacial lake floods.

As a poor nation without even its own helicopter, Bhutan lacks the resources to combat global warming. It is carrying out the work at Thorthormi glacier with the help of money from various international donors, including US\$3.5 million from the Least Developed Countries Fund, created under the United Nations Framework Convention on Climate Change. The global cost of adaptation could total hundreds of billions of dollars a year — orders of magnitude more than what is available to poor countries at the moment. During December's UN talks in Copenhagen, developing countries will be pushing for more

generous — and reliable — funding to help them mitigate the impacts of climate change. As the first nation to get adaptation money from the Least Developed Countries Fund (see 'The long wait for adaptation money', page 1045), Bhutan is something of a pioneer among developing nations in their quest to adapt to a warmer future. And the struggles at Thorthormi glacier illustrate the enormous obstacles that adaptation efforts still face.

Dangerous dam

The sounds of global warming are deafening at Thorthormi glacier. Every few minutes, a block of ice rips off the glacier and crashes into the lake in a trail of dust and ice. These are some of the tallest mountains in the world and form Bhutan's northern boundary with Tibet.

To get a look at the hazard posed by the melting glacier, Karma Toeb, the project's glaciologist and team leader, scrambles to the top of a moraine — a steep ridge of loose, angular boulders built up by the glacier as it pushes debris along its edges. This moraine is a dam between two bodies of water (see map, overleaf). To the east, Toeb points out how several slushy grey-brown ponds have formed



CRYOSPHERE RES. LAB., GRADUATE SCHOOL ENVIRON. STUD., NAGOYA UNIV.

on top of the Thorthormi glacier. To the west lies a much larger lake (pictured above) called Rapstreng — vast and milky green — about 80 metres lower in elevation. Toeb first came to this site in 1997, during a project to shrink Rapstreng lake. Over three years, a thousand labourers widened and deepened Rapstreng's natural outlet to lower the lake by 4 metres.

At the time, nobody was worried about the Thorthormi glacier next door. "Thorthormi was almost pure ice," says Toeb. "We could cross over and walk on the glacier to do our research." The glacial lake didn't even appear in an exhaustive study of the region's dangerous lakes in 2001 by the International Center for Integrated Mountain Development (ICIMOD), based in Kathmandu, Nepal^{2,3}.

It is only within the past decade that researchers realized that Thorthormi could pose a threat. Thorthormi's ponds were expanding and merging to form larger bodies of water. The changes have been dramatic even in the past few months. "Just before we started our work here in July this year, that part of the lake was water," says Toeb, pointing down to a number of icebergs. "The ice blocks have been breaking off the mother glacier upstream."

But the rate at which Thorthormi glacier is melting is not the only concern, says Toeb, motioning at the boulders underfoot. The moraine separating the Thorthormi and Rapstreng lakes is at places only around 30 metres wide and is prone to landslides. Geophysical testing in 2008 showed that the moraine contains a substantial amount of ice. As that ice melts, along with the rest of the glacier, the ridge could collapse, releasing the water

from Thorthormi into Rapstreng below. Then there would be a combined outburst flood, Toeb says.

Building pressure

A four-year study led by Hermann Häusler from the University of Vienna in collaboration with Bhutan's geology and mines department predicted the moraine between Thorthormi and Rapstreng could give way as early as 2010, because of the hydrostatic pressure of the growing Thorthormi glacial lake. A combined flood would unleash around 53 million cubic metres of water down the Pho river⁴.

Every country within the Himalayan region has suffered a glacial outburst flood at some point³, and Bhutan is no exception. A number of glaciers in that country are missing sections of their moraines, suggesting that lakes have burst there in the past. Half a century ago, floods originated from the Lunana region in northern Bhutan, which is home to Thorthormi and many other glaciers. The floods may have come from glacial lakes, but there were no studies of them at the time. The only scientifically documented glacial flood in Bhutan's history came from a lake at the bottom of Luggye glacier, on Thorthormi's eastern margin, 15 years ago.

Ponds first appeared on the Luggye glacier in the late 1960s, and by the early 1990s the glacier was retreating by up to 160 metres a year². On 7 October 1994, the glacier's moraine broke and released an estimated 18 million cubic metres of water and debris down the Pho river,

killing 21 people and razing fields and settlements downstream with a jumble of uprooted trees, boulders and mud.

Memories of that flood still haunt Dawa Gyeltshen, a herder who lives in the village of Taksho, beside the Pho river around 20 kilometres downstream of Luggye. He remembers being woken by a thunderous noise early that morning. For hours he raced around outside his house in the dark, terrified as the river rose.

It wasn't until dawn that Gyeltshen saw the extent of the damage. The forested river valley and a large portion of the fields in front of his house had been washed away and in their place were mounds of boulders and glacial silt. Today, the landscape is

recovering. Pockets of shrubs and wildflowers have settled in along the river and Gyeltshen's remaining land is golden with this year's crop of buckwheat. White prayer flags line the rugged edge of the property, marking the course of the water during the flood. But even now, Gyeltshen says he still has trouble sleeping through the night, worried about another flood.

A combined flood from Thorthormi and Rapstreng lakes could cause at least ten times the damage and fatalities of the 1994 event⁴. In the past 15 years, hospitals and schools have popped up along the river, along with a partially completed \$760-million, 1,200-megawatt hydropower dam project.

It was in part because of the 1994 disaster that Bhutan became a leader in adapting to climate change. Following the flood, scientists from Bhutan, Japan and Austria, among other





Prayer flags mark the course of the water during a flood 15 years ago.

countries, launched research programmes into the glaciers in the Lunana region. When the Least Developed Countries Fund was created in 2001, “we had a lot of good information on glacial floods”, says Thinley Namgyel, the deputy chief environment officer at the National Environment Commission in Thimphu.

Gross national happiness

Bhutan’s work on adaptation is backed by a strong record of environmental protection. As a follower of the Buddhist tenet of non-extremism — the Middle Path — Bhutan’s monarchy has stressed that modernization and economic development are important, but not at the expense of the country’s natural environment and cultural traditions. In 1972, Bhutan’s fourth King, Jigme Singye Wangchuck turned away from the classic yardsticks of growth, such as gross domestic product, to focus on a more holistic approach to his people’s well-being. The measurement, which he called gross national happiness, is based on several measures of environmental and social contentment.

The country has some of the most progressive — and controversial — environmental regulations in the world, including bans on plastic bags, timber exports, hunting and even tobacco sales. They’ve also minimized tourist traffic by charging visitors to the country around \$200 a day.

The nation’s proven environmental record, coupled with the immediacy of the risks of a glacial flood, made Bhutan’s adaptation needs especially attractive to donors. “From the start,

Bhutan was in the driving seat,” says Bonizella Biagini, a senior programme manager at the Global Environmental Facility, which manages the Least Developed Countries Fund. Other developing nations have been less successful in getting adaptation money through that fund. “If everyone behaved like Bhutan, every project would be in the process of being implemented,” says Biagini.

Bhutan has also received \$1.3 million in adaptation support from other sources, including the Austrian Development Cooperation,

of the lake, the extreme altitude, the boulder-strewn terrain and the unpredictable weather have made the project logistically difficult.

The project’s engineer, Karma Tenzin, had hoped originally to fly a couple of excavators to the site and do the digging in a few weeks. But soon after visiting Lunana this year, he realized an industrial solution would be difficult. The nearest potential helicopter landing was an hour and a half away by foot. Even if an excavator was brought in, it could topple on the uneven terrain, says Tenzin. “Everywhere you look are boulders,” he says. “Bringing in an excavator is useless.”

The price was also a factor: without its own helicopter, Bhutan would have had to hire one from Nepal for several thousand dollars per trip. Because unemployment is rising in Bhutan, the project board reasoned that it would be better to pump the money into the country’s economy — by paying local horsemen to transport the project’s goods and by providing excavation jobs. So in July, in the middle of the monsoon season, hundreds of workers set off on the 9-day journey to Lunana from the capital. At the site, using shovels and spades, they load cobbles into strips of burlap and carry them to the sides of the channels. Boulders are broken crudely with blunt hammers and hauled off-site with weathered rope. At the end of September, a few jackhammers and chisels arrive to help split boulders. A shipment of rope also arrives, although Tenzin admits that it is unlikely to last long. “At least for the next two weeks we will be able to pull rocks,” he says. “After that, the rope is likely to be all broken again.”



the conservation group WWF Bhutan and the United Nations Development Programme.

Yet even with money and political will-power, the task is daunting. The main goal at Thorthormi glacier is to lower the lake’s water level by 5 metres, which would reduce the hydrostatic pressure pushing on Thorthormi’s unstable natural dam.

Under normal conditions, the lowering would be an easy task, but the remote location

A. NAYAR

JAXA

The working season for the project is less than four months a year, because snow blocks the path to the site for all but July to October. But heavy rainfall this year washed away several key bridges, delaying work for an additional month. "We almost cancelled the project for this year," says Dowchu Dukpa, the project's manager. "We thought: by the time we get there, we will have to go back."

But the project went ahead. When the last workers put down their tools for the year on 15 October, they had lowered the lake by around 90 centimetres. This was considerably less progress than the original goal of 1.67 metres for this first year, but the group thinks it can make up the difference during the next two years of the project. "I am sure the goal will be achieved," says Tenzin.

Practical challenges

Even if the final goal of 5 metres is achieved by 2011, glacial experts cannot say how much that will reduce the risk of a moraine burst. Part of the problem is the lack of basic information about the Thorthormi glacier and its lake.

The glacier is covered in boulders and silt, which makes it difficult to measure how quickly the ice is retreating, says Koji Fujita, a glaciologist from the University of Nagoya in Japan.

The lake's drifting icebergs also make a comprehensive bathymetric survey extremely dangerous, says Fujita. Bhutanese officials only took a few depth measurements in 2008 with a simple string and weight, because strong winds made it difficult to handle the boat. "We don't know how much water will come out of the lake or how fast it will flow in an outburst," says Fujita.

What's more, developing countries such as Bhutan don't have the technical and financial resources to study the changes that are happening in their countries. High-resolution satellite imagery can be expensive and the country can afford it only infrequently. Difficult and remote terrain also hampers their monitoring work. Every scientist and bit of instrumentation has to make the journey to Lunana on foot. To top it off, they are contending with a landscape that is rapidly changing because of global warming.

Project leaders reason that their strategy at Thorthormi must reduce the hazard there to some degree. "If the lake is lowered by some metres, that much pressure is released from the moraine wall," says Toeb. "So the chance of the moraine wall failure is less."

But Toeb admits that the work will not eliminate the possibility of a flood. "A piece of ice could detach from the mother glacier, fall into the lake and generate a surge wave," says Toeb. "Or if there is a big seismic event in this



The long wait for adaptation money

The cruelty of climate change is that developing nations are the most vulnerable and least able to afford protective measures. To help them cover the billions of dollars it will cost to adapt, the United Nations Framework Convention on Climate Change set up funding mechanisms during climate talks in 2001.

The main source of money, the Least Developed Countries Fund (LDCF), is supported by donations from rich countries and currently has US\$176 million in pledges. The fund is managed by the Global Environmental Facility (GEF). To qualify for money, countries first need to itemize and prioritize their urgent needs in a document called the National Adaptation Programmes of Action (NAPA).

Although 42 countries have submitted their NAPA documents to date, only one country, Bhutan, has physically started implementing its projects. Developing countries blame this delay on the GEF's "bureaucratic red tape".

The situation is improving; the GEF has endorsed projects in eight more countries so far this year. But there is another bottleneck: There isn't enough money in the LDCF to go around. Each developing country is limited to \$5 million, enough for only a fraction of their NAPA priorities. To implement all NAPA projects currently submitted, the fund would need around \$1.7 billion

— nearly ten times what it has been pledged by developed nations so far.

Bhutan is moving ahead with three of its nine NAPA priorities: capacity building for disaster management, lowering a glacial lake (shown above) and an early warning system for glacial floods. But the country is anxious to start on more of its priorities, says Thinley Namgyel, deputy chief environment officer at the National Environment Commission in Thimphu, who helped coordinate the country's adaptation plan. "We were told we had to wait until the LDCF is replenished," he says.

The Himalayan nation was the first to receive funding in part because it moved much faster than others. Bhutan was the first country to submit its fully completed NAPA, in 2006. It also lacks the degree of poverty and corruption that troubles many developing nations.

Most other developing nations have struggled with the process. Because the nations have less capacity to plan and implement adaptation projects, the GEF requires them to work through an implementing agency, such as the UN Development Programme (UNDP), the UN Environment Programme or the World Bank.

This architecture adds bureaucratic layers that slow the process. Although Bhutan's neighbour, Nepal,

initiated its NAPA in 2002, the country still has not completed its initial prioritized list. Batu Uprety, joint secretary for the ministry of forests and soil conservation in Kathmandu, says Nepal's implementing agency, the UNDP, was more of a barrier than a facilitator.

Bonizella Biagini, who oversees the LDCF, says that reviews of NAPAs take at most 15 days. "When a project is submitted to us, we have to look to see if it meets the LDCF criteria," she says. "If it doesn't, I send it back. Sometimes it comes back in two weeks, sometimes it comes back in a year, sometimes the government folds."

Last December, during the last round of UN climate negotiations in Poznań, Poland, several developing countries complained that there was too much red tape involved in accessing GEF funds. They called for direct access to GEF funding. "If a country wants to work through an implementation agency that is fine," says Uprety. "But if a country doesn't want to, that shouldn't be a barrier."

It isn't clear whether the LDCF will survive in the next climate agreement, as developing countries look for easier ways to access money. A possible funnel for adaptation money in the next climate agreement is the Adaptation Fund, which was also set up in 2001 but has yet to accumulate much money or to distribute any funds. **A.N.**

A. NAYAR



Karma Tenzin uses a stick to measure the level of Thorthomi lake.

area, the whole moraine wall surrounding the glacial lake may collapse.”

Because artificially lowering Thorthormi isn't enough to prevent a flood, the government intends to install an automatic flood early warning system in the Pho river valley, which will also be paid for with money from the development fund.

The Bhutanese Department of Energy currently runs a manual warning system in the Lunana region. A couple of attendants equipped with a VHF radio and satellite phone are supposed to monitor the lakes and river three times a week. But the system isn't working. “We never see them at the site,” says Toeb, who has reported the problem to the department.

The new automatic system will record the water levels using sensors. In the case of an outburst, the system would send out a signal to towers downstream to warn communities of an impending flood, potentially providing hours of advance notice.

Because Bhutan does not have the capacity or expertise to do the work on its own, its Department of Energy has put the project out to tender, in hopes of an international bid within its budget. Construction is expected to start next year, says Karma Chhophel, the head of the department's hydrological- and meteorological-services division.

The entire Thorthormi mitigation project will last at least four years and cost \$7.4 million. But Thorthormi is only one of thousands of glacial lakes in Bhutan. ICIMOD has identified 25 potentially dangerous lakes in Bhutan, but the list was based on crude satellite data².

The centre will release an updated inventory of dangerous lakes in Bhutan next year based on satellite imagery with “better spectral, spatial and temporal resolution,” says Pradeep Mool, a remote sensing specialist with ICIMOD, who was involved in the study. The inventory will also rank lakes based on socioeconomic information, including the potential loss of

life and damage to infrastructure. “All the lakes need attention,” he says. “But we have to prioritize where we do the big-scale mitigation work, the early-warning system, and the hazard zonation maps.”

Spy data

But qualitative analysis of satellite images isn't enough to determine which lakes are dangerous, according to Fujita. He is trying to develop quantifiable criteria, such as the angle between the level of the water in the glacial lake and the slope of the moraine, that dictate a moraine's stability and the likelihood of a glacial outburst. His team is testing that hypothesis by comparing new satellite images with data from old American spy satellites to create digital elevation models of lakes that burst in the past. The researchers will also develop their own list of the region's dangerous lakes next year, but the results are likely to be substantially different from ICIMODs, says Fujita. “Some of its 25 dangerous lakes are not a hazard, but we found more that ICIMOD did not point out,” he says.

Regardless of which lakes are a risk now, the number will rise in the future, as glaciers continue to melt. “Thorthormi is a very good lesson,” says Toeb. “Nobody knows what will happen, taking into account all the changes in climate; the same situation may happen to any of the glaciers in Bhutan.”

Although glacial lake bursts cause considerable damage in the Himalayas, an even bigger catastrophe would come from the disappearance of those very same glaciers, and the water they produce, which is predicted within the next few decades. That loss could significantly harm the 69% of the Bhutanese population that relies on farming, mostly subsistence. A decrease in water flow could also seriously

affect the country's plan to boost its hydro-power production by 10 gigawatts by 2020. “We don't know the medium- or long-term costs of climate change in Bhutan,” says Namgyel. “A few decades down the line, the glaciers will retreat and we are not sure what impact it will have on the economy.”

The impacts will reach far beyond Bhutan's borders. The glacier-fed rivers that flow south from the Himalayas are the arteries of south Asia. It is estimated that the retreat of glaciers will affect the water supply of roughly 750 million people across South Asia and China, says Rajendra Pachauri, the chairman of the Intergovernmental Panel on Climate Change.

Across Asia, there are countless cases like Thorthormi, where the needs are great and the resources scarce. Regarding the effects of climate change and their costs, “every single estimate that people have come up with has been exceeded by reality,” says Pachauri. “The impacts of climate change are clearly turning out to be much worse than what we had anticipated earlier.”

Standing on Thorthormi's natural dam, the scale of the problems ahead strain the imagination. From here, the sounds of the labourers struggling 100 metres below fade to whispers

on the wind. At this distance, it is impossible to see their frost-bitten cheeks, or the cuts and bruises earned during their months of labour. In a line of 30 people, they look like a caterpillar pulling pebbles across a path. They are dwarfed by the size of the lake, the glaciers above, and

the mountain of work confronting Bhutan and the rest of the world, as it tries to keep pace with the changing climate.

Anjali Nayar is an International Development Research Centre fellow at Nature.

“All the lakes need attention. But we have to prioritize where we do the big-scale mitigation work.”

— Pradeep Mool

1. Cruz, R. V. et al. *Asia. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) 469–506 (Cambridge Univ. Press, 2007).
2. Bajracharya, S. R., Mool, P. K. & Shrestha, B. *Impact of Climate Change on Himalayan Glacier and Glacial Lakes: Case Studies on GLOF and Associated Hazards in Nepal and Bhutan* (ICIMOD, 2007).
3. Mool, P. K. et al. *Inventory of Glaciers, Glacial Lakes, and Glacial Lake Outburst Floods: Monitoring and Early Warning Systems in the Hindu Kush-Himalayan Region, Bhutan* (ICIMOD, 2001).
4. Brauner, M., Leber, D. & Hausler, H. *Glacier Lake Outburst Flood (GLOF) Mitigation Project, Lunana, Bhutan, Technical Mitigation Measures Thorthormi Outlet*. (Department of Geological Sciences, University of Vienna, 2003).

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen. Watch a video at go.nature.com/VGGayN.

Counting carbon in the Amazon

If the next climate treaty tackles deforestation, tropical nations will need to monitor the biomass of their forests. One ecologist has worked out a way to do that from the sky, finds **Jeff Tollefson**.

Greg Asner peers out an open window, taking stock of the jungle as the single-engine prop plane chugs over a pair of scarlet macaws gliding among the treetops 120 metres below. The Peruvian Amazon stretches in all directions, painted in countless shades of green, accented here and there by patches of purple, pink and yellow. Occasionally, naked white trunks rise amid the leaves, a reminder that even the rainforest has deciduous tendencies.

Forty-five minutes into the flight, Asner spots his quarry: narrow red trails, barely visible, then a fallen tree in the middle of an otherwise intact canopy. The cause isn't immediately clear to the untrained eye, but Asner knows all too well. "When trees die in the tropics, they don't just fall over," he says as the plane passes over more downed trees, a road, then a small clearing that contains stacked logs and a bulldozer. It is a legal concession, authorized by the Peruvian



government to extract just three species of hardwood trees. As the plane veers away from the clearing, Asner gives his verdict. "The biomass levels are going to be a lot lower here," he says, "but it really is low-impact logging compared with the mayhem of Brazil."

As a tropical ecologist with the Carnegie Institution for Science's global ecology department in Stanford, California, Asner has developed a keen ability to interpret the rainforest from great heights. Frequently operating with oxygen masks at high altitude, his team uses a powerful laser system to map trees and calculate the biomass of the forest. Satellites extend his view across the tropics, and he has developed automated software that can track annual changes in forest cover and calculate the biomass of the vegetation. The system can even spot small logging operations like the one he just passed, which escape detection in most satellite studies.

The fully integrated system is designed to measure the

G. ASNER



amount of carbon locked up in forests and to track changes over time — an exercise that may become a crucial foundation of the new climate treaty that global leaders are hoping to sign at the United Nations Climate Change Conference in Copenhagen this December. Tropical deforestation accounts for up to 20% of the carbon dioxide emitted by humanity each year and there is broad agreement on the need to include a forest-protection element in the new treaty. This component — known as REDD, for Reducing Emissions from Deforestation and Forest Degradation — would allow developed nations to meet their required emissions-reduction targets in part by paying tropical countries to preserve their forests, which keeps carbon in trees and out of the atmosphere.

But first, tropical nations must determine how much carbon is in the forest, a notoriously difficult task. A leading researcher on remote sensing in the tropics, Asner is out to prove that developing countries can quickly and cheaply perform their own analyses, then move on to long-term carbon monitoring. He has come to Peru to demonstrate the technology and he will present the results of this proof-of-concept test at a REDD meeting he is planning to coincide with the negotiations in Copenhagen. Success in Peru, he hopes, will bolster efforts to include a strong forest carbon component in the agreement.

The European Union last year called for a halving of deforestation by 2020, a goal that has since picked up political momentum. Estimates range widely, but reaching that target could mean pumping some US\$20 billion into tropical countries each year, according to the Union of Concerned Scientists in Cambridge, Massachusetts. Venture capitalists see profits in forest carbon and have approached Asner with business offers. Instead he has licensed his technology, dubbed CLASLite, for Carnegie Landsat Analysis System Lite, and is providing it for free to governments and others. Starting in Latin America, his team is training scientists, officials and advocacy groups on how to use the software. He likes to say that he is putting himself and other scientists out of business by injecting a decade's worth of work into the public sphere.

All this effort in testing and training has meant devoting less time to academic research, but Asner says that his work on REDD has been nothing short of rejuvenating. "This is more fun than anything I've ever done as a scientist," he says. "These forests are really special. We need to get these people some cash and protect them."

Sweeping the forest

For Asner's crew in Peru, work starts early each morning. Today, a three-man team heads to a small airport in the southeastern part of the country before sunrise in an effort to get as much mapping done as possible before the jungle pumps enough water into the air to form midday clouds. Ty Kennedy-Bowdoin runs the Light Detection and Ranging system, or LIDAR, which sweeps a laser back and forth, blasting the forest with 70,000 laser pulses per second. A sensor continuously records the signals as they bounce off leaves, branches and other objects. These data enable the researchers to calculate the height, structure and density of the forest, and they use this information to determine how much biomass it holds.

Beside Kennedy-Bowdoin in the plane is James Jacobson,



J. TOLLEFSON

Greg Asner is trying to help tropical countries keep their forests intact.

"The forces against deforestation are getting stronger, but the forces driving deforestation are getting stronger too."

— George Powell

who monitors a hyperspectral imager that takes pictures using frequencies of light that range from the visible into the infrared. The spectral data are used mostly for research on forest biodiversity, although they can also be combined with the LIDAR data to create full-colour three-dimensional images. On this particular flight, Kennedy-Bowdoin eyes a computer monitor to confirm that the LIDAR readings are coming in properly as the plane slowly flies back and forth. But after several passes small clouds move in and block the laser. The plane returns to the airport, allowing the crew to break for lunch before another run in the evening, once the clouds have dissipated.

Back on the ground, Asner and a few colleagues head off into the jungle. They drive west along the InterOceanic Highway that crosses South America, then take a small logging road into the rainforest. It's dark by the time they finish a final hike to the Tambopata River and hop on a motor boat for a 45-minute ride to a research station upstream; a spotter scans the river with a flashlight, occasionally directing the boat away from logs.

There they meet up with a second team that is organizing ground plots. After waking up to the sound of monkeys in the morning, the team divides into several groups and heads out into the field; they have to finish around 30 plots before moving on to other locations throughout a study area larger than Denmark.

At one of the field plots, strings stretch 30 metres out from a pole in multiple directions. In the middle, the researchers are busy running a tape measure around any tree bigger than 10 centimetres in diameter; in some places, they sample down to 5 centimetres. A laser finder gauges the height of each tree. With those data, Asner's team can calculate biomass for every tree in the plot, and those numbers are then plugged into equations to calculate the biomass for the patch of forest.

With a smile, Asner explains that this cumbersome process is currently the gold standard for biomass assessment. "People running tape measures around trees. This is

what we've got to get away from," he says.

Asner does this by taking to the skies. The laser pulses of the airborne LIDAR provide the main biomass estimate for large swathes of forest. The ground plots help his team to interpret and verify the laser readings for different types of vegetation (see 'How to measure a forest'). Old-growth forest filled with hardwood, for example, contains much more biomass per hectare than do regions dominated by bamboo. The CLASLite system then adds the numbers up according to vegetation type and extent to produce an estimate for the entire forest. Initial results suggest that these remote-sensing techniques are just as accurate as plots, Asner says, but they allow a small team to cover vast territories in a short time.

"Greg has already demonstrated that a tiny group of people can deploy this system over the scale of an entire country; the preliminary activities in Peru make that clear," says Chris Field, Asner's boss at Carnegie who is co-chair of the impacts, adaptation and vulnerability working group for the Intergovernmental Panel on Climate Change (IPCC). "I think it's an incredibly important development in science, and I'm a tremendous fan of his ability to make these things happen."

Sandra Brown, a leading biomass expert at the non-profit organization Winrock International in Arlington, Virginia, acknowledges that plot-based assessments, although reliable, simply cannot cover as much territory as remote-sensing methods. She also says that published biomass estimates vary wildly for a given region, highlighting the need to establish baselines across the tropics that everybody can agree on. Stopping short of a blanket endorsement, she says that she is encouraged by Asner's initial results and is anxious to see the process in action. "I think it's got a lot of promise," she says.

The current project in Peru costs around \$430,000, half of which was funded by the Norwegian government through a grant to the environmental group WWF. That equates to roughly 10 cents per hectare, with about half of the money spent on fuel and the plane, but Asner expects the costs to



fall over time. Countries could elect to develop their own LIDAR capacity, he says, or hire one of the more than 100 commercial operators around the world. For perspective, Asner estimates that he could map all of the world's tropical forests with his system for roughly \$15 million–\$20 million.

Map quest

Asner owes his interest in forests to a hurricane named Iniki, which swept the Hawaiian islands in 1992. Asner was stationed there with the US Navy at the time, and he wound up surveying forest damage on Kaua'i after taking a job with the Nature Conservancy the following year. Asner recalls being "enthralled with the beauty, the wildness, the enormous hurricane damage and the onslaught of invasive species". His first scientific paper focused on the carnage caused by Iniki, but it was his work with invasive species that made him realize he was missing something as he struggled to assess the forest: maps.

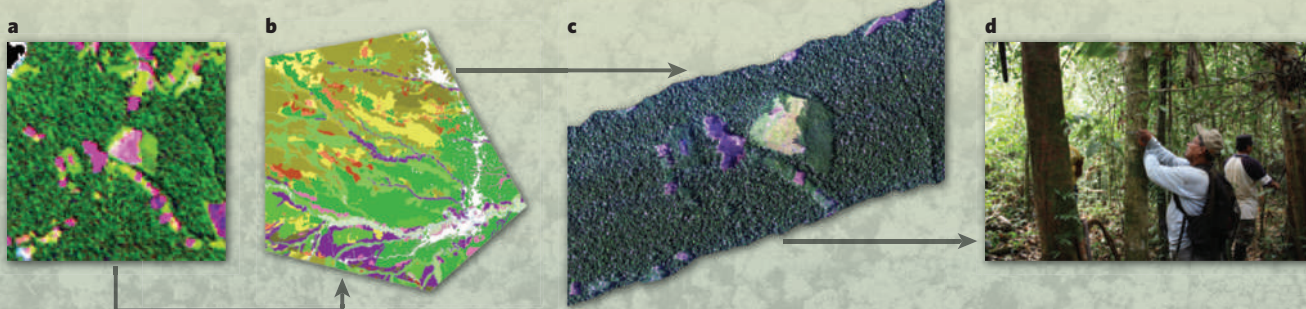
His quest for spatial data took him to the University of Colorado at Boulder for graduate work in ecology, biogeochemistry and remote sensing, and then all over the world. For Asner, now 41, it is hard to separate research from the rest of his life. His wife, Robin Martin, is a postdoc in his lab and they spend most of their time together in the field. And for the first test of the new biomass system earlier this year, he chose to measure Hawai'i, where he lives part time.

In that analysis, Asner's group identified roughly 48 million tonnes of above-ground biomass across the island. That is about 40% lower than the results the researchers obtained from published data and a simple protocol established by the IPCC, which would be the probable starting point for any country that is setting up a carbon-assessment programme. Asner's method has cut the calculated uncertainty in half and he says his system is accurate enough to meet criteria set out by the IPCC for advanced biomass monitoring (G. P. Asner *Environ. Res. Lett.* **4**, 034009; 2009). Satisfied with his results in Hawaii, Asner set his sights on Peru, where he has worked for several years on estimating biomass and other research.

"People running tape measures around trees. This is what we've got to get away from."

— Greg Asner

HOW TO MEASURE A FOREST



Greg Asner's automated software, called CLASLite, analyses raw satellite imagery to produce a current picture of forest cover and disturbance (a). These data are then lined up against existing maps that categorize different types of vegetation (b). Laser-based instruments called LIDARs are flown on planes to measure the density of biomass in each vegetation type (c). A small number of ground plots (d)

help to calibrate and validate the LIDAR estimates. CLASLite then applies the biomass estimates across the entire study area to produce an overall biomass estimate, which can then yield the total carbon content of the vegetation. After an initial assessment, CLASLite can monitor biomass levels on an annual basis; countries could deploy LIDAR occasionally to assess areas of particular interest.

G. ASNER



Laser-based measurements assess forest biomass.

The Carnegie crew began its project in Peru with a CLASLite analysis of NASA Landsat images around Puerto Maldonado, which is tucked into a corner of the southeastern Peruvian Amazon near Bolivia and Brazil (see map, opposite). Puerto is a poor frontier town of nearly 30,000 inhabitants, with more descending from the Andes along the newly paved InterOceanic Highway every day in pursuit of riches provided by logging, agriculture and, most recently, gold mining. Rainforest is nowhere to be seen in this dusty Amazonian town, but thousands of motorcycles buzz about, many powering a local variation on the rickshaw taxi.

The region around Puerto Maldonado, however, is one of the most biodiverse locations on the planet and a hub of deforestation. This makes it a prime target for Asner, who says that the government has shown it is ready to protect its forests.

In 2008, Peru created its first environment ministry. Vanessa Vereau, former vice-minister, says that Peru is working to build up regional governance even as it increases federal enforcement capacity by putting more police on the ground.

It is also installing federal prosecutors in rural areas to ensure that everybody — government officials, companies and individuals — face real consequences when they break environmental laws.

Success is by no means assured, but the government has not even tried to impose such order until now, says Vereau. Peru has also established its own version of a REDD programme by creating monetary incentives for indigenous communities that agree to protect their forests instead of cutting them down.

Unlike Brazil, where squatters take advantage of undesignated territories, Peru has designated its land in the Amazon as parks and conservation areas as well as concessions for timber, Brazil nuts and other resources. Along the roads outside Puerto, however, are mostly fallowed fields in place of former rainforest, illustrating how poverty and poor soils combine to create endless pressure to clear land. Although the new federal actions are encouraging, governance remains weak and the fundamental frontier economics have not changed, says George Powell, a wildlife biologist with the WWF who is based in Peru and has become one of Asner's partners on the project.

"The forces against deforestation are getting stronger, but the forces driving deforestation are getting stronger too," Powell says one evening after the team returns from the

airport to a small hotel that serves as base camp for Asner's team and for local bosses who are directing work on the InterOceanic Highway. Even as the Peruvian government paves the way for bigger trucks to haul more resources out of the jungle, he says, Asner is blazing a trail that could make carbon payments a reality and help tilt the balance in favour of standing forests. "Most people are still talking about what remote sensing can and can't do," says Powell. "Greg is already down the road putting it in the hands of users. He's three steps ahead."

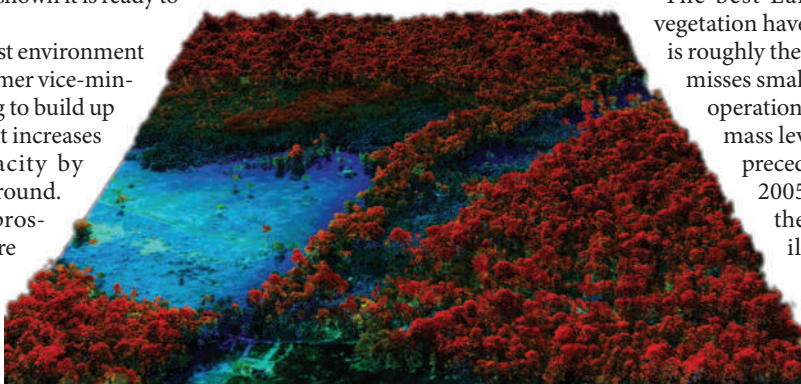
Seen from the vantage of a satellite, the deforestation around Puerto extends like fish bones, following roads in long, crossing lines. These are the obvious clear cuts, where forest has been chopped down for farming. Asner's CLASLite analysis also picks up countless blips of small-scale logging far out into the forest.

The best Landsat images for analysing vegetation have a 30-metre resolution, which is roughly the size of a large tree crown. This misses small roads and hardwood logging operations that substantially reduce biomass levels deep within the forest and precede widespread deforestation. In 2005, Asner published a paper on the Brazilian Amazon in *Science* illustrating how probabilistic algorithms could be used to sniff out the hidden spectral signals of selective — and often illegal — logging (G. P. Asner *Science* **310**, 480–482; 2005).

That study also roughly doubled previous estimates of the amount of forest affected by human activities, while increasing the estimated greenhouse-gas emissions from the Amazon by up to 25% compared with deforestation alone. It also made a name for Asner and helped him to attract financial backing from the MacArthur Foundation, the Gordon and Betty Moore Foundation and other organizations. In total, they have given him more than \$11 million to support his biomass and deforestation work, as well as other biodiversity research.

Some of that funding has gone into producing and testing a new version of CLASLite, which is designed to make it easy to conduct a biomass assessment and to set up a monitoring programme. Import a publicly available satellite image and CLASLite will correct for atmospheric conditions at the time the image was taken, then analyse the spectrum of each pixel. Vegetation that photosynthesizes has a different spectral signal from dead trees, rocks or soil. A 'Monte Carlo' analysis then produces a range of possible combinations that converge on the most likely explanation for the data. For instance, bare soil is rarely exposed to the sky in a

G. ASNER



Laser readings combined with other data yield a three-dimensional colour image.

J. TOLLEFSON/G. ASNER



The CLASLite software developed by Greg Asner reveals where new forest destruction (green and yellow specks) extends into a protected area (red).

fast-growing rainforest; if CLASLite picks up even a tiny red signal associated with the region's iron-rich soils, and that red colour extends in a line through multiple pixels, the most likely explanation would be a road.

In addition to analysing the satellite data, the software automatically pulls in existing vegetation maps and suggests locations for both aerial measurements and ground plots. In other words, the program helps to plan an integrated biomass analysis similar to the one that Asner is conducting around Puerto. "We were trying to figure out what users want, and then we finally realized they want everything, and they want to be able to hit a button," Asner says. "There's 10 years wrapped up in this little widget. That's super-secret sauce."

As of this week, Asner has trained more than 240 people in six countries on the software, including several Peruvian government officials. For Peru, the project represents an opportunity to build up its scientific capacity and perhaps even leapfrog Brazil, which currently has the world's most advanced forest-monitoring programme. Deforestation is responsible for a large — and unknown — fraction of emissions in Peru, and the country is banking on REDD as a new development aid to buttress conservation efforts while reducing carbon emissions.

"It's a new issue for us, but there is a real political will to protect the forests," says Vereau. She sees the current partnership with Asner as a pilot project that could go national "if it has strong results".

Asner acknowledges the tremendous challenges ahead. He must convince not only the Peruvian government that the system works but also the broader scientific community and ultimately, perhaps, policy-makers who are debating how to restructure the global economy around carbon. But if he has any lingering doubts, they are smothered by his boundless enthusiasm. Flying over the Amazon, Asner can't help but marvel at the sights below. "Wow, there's another

super-giant! That's a 300-year-old tree," he says. "These are just amazing organisms!"

Asner's work could soon have a much broader impact thanks to an agreement he recently reached with Google.org on a forest-monitoring application that would be freely available on the web. That partnership, which could be announced as early as next month, could further reduce start-up costs for tropical countries by providing them with processing power and easy access to freely available satellite data from agencies such as NASA and the Brazilian Space Agency. Asner isn't allowed to talk about it, and Google officials won't go into detail except to say that they are building a rainforest-monitoring platform and that Asner's CLASLite software will be part of the package.

Dan Nepstad, a tropical ecologist at the Woods Hole Research Center in Falmouth, Massachusetts, predicts that the remote-sensing community is on the verge of a major transformation that will open the doors to high-quality forest monitoring on a global scale. He cites progress on many fronts but says that Asner has shown a particular knack for making his science relevant and useful to policy-makers, particularly those in tropical countries. "There has to be trust in the forest-monitoring data, and these nations have to see them as their own," he says. "There's this face-to-face collaboration that is really critical."

Having spent most of his time over the past two years working to deploy his vision of REDD, Asner is now planning to shift into other projects. He is already working on an instrument for his biodiversity and biomass research that combines a more powerful LIDAR with sensors that capture reflected light in 440 frequencies. He is also taking another look at the issue of selective logging.

Asner recently completed his first, as yet unpublished, analysis of logging across the tropics and found that small-scale logging operations have a footprint that is 20 times larger than the more obvious wholesale deforestation. Right now everybody is focused on deforestation, but degradation raises a new set of scientific questions, including how much carbon is being lost and how quickly the forests will recover. Asner calls this a "future frontier science activity", which means he'll tackle the issue next year.

First, he has to finish his work in Peru. Walking along the logging road back from the Tambopata River, Asner comes across a series of trees that have recently been cut down and cast aside. They were old, hollow and unusable, but the loggers only discovered that fact after felling the trees. They could have checked the wood without hurting the trees, says Asner, by inserting a chainsaw directly into their trunks and then pulling it back out. Now the wood is simply rotting away, its carbon needlessly committed to the atmosphere.

"There are millions of trees just like this," Asner says, kicking one of the stumps. "Can you hear them falling? I can. I can hear them in my sleep."

Jeff Tollefson is a reporter for *Nature* based in Washington DC.

"There has to be trust in the forest-monitoring data."

— Dan Nepstad

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen.

CORRESPONDENCE

Sharing: project will make climate data freely available to all

SIR — Some of the concerns expressed in your News story on the difficulties of collecting and sharing climate data across countries are unjustified (*Nature* **461**, 159; 2009).

The World Meteorological Organization's task force is helping to develop a Global Framework for Climate Services to link weather predictions, projections and information with climate-risk management and adaptation. This international service will provide free and unrestricted collection and exchange of meteorological data.

Because the information is for the public good, there will be no competition or exclusion in accessing it. Application by one user will not reduce its availability to others. It would also be impossible — or very costly — to exclude potential users from using the data for their own benefit.

Another advantage is that although climate information is expensive to produce, it is relatively cheap to reproduce and distribute — making it economically efficient to supply these valuable data to all for free.

Don Gunasekera Centre for Complex Systems Science, CSIRO Marine and Atmospheric Research, Commonwealth Scientific and Industrial Research Organisation, GPO Box 3023, Canberra, ACT 2601, Australia
e-mail: don.gunasekera@csiro.au

Sharing: public databases combat mistrust and secrecy

SIR — You cite compelling reasons for the scientific community to share data (*Nature* **461**, 145 and 168–173; 2009). But there is also a case for extending this to the broader community, including the general public.

Vertebrate palaeontology is a particularly rich candidate in

this respect, as evidenced by blogging activity and busy internet traffic. Requests by amateur enthusiasts for copies of publications, measurements and photographs of fossils are commonplace.

To harness some of this enthusiasm, we launched the Open Dinosaur Project (<http://opendino.wordpress.com>) last month. Participants include students, professional scientists and artists, who enter measurements of fossil specimens from the literature and personal observation into a central public database. Participants may also contribute their expertise in data analysis and interpretation. All contributors will eventually be listed as authors on the first publication arising from the database.

Many older papers include data-rich tables of measurements, which are essential for comparisons between specimens. But there is a recent disturbing tendency to omit such information, even in studies that analyse hundreds of measurements for documenting evolutionary trends — perhaps because authors believe their exclusive access to the raw data gives them a competitive edge (*Nature* **461**, 160–163; 2009). The Open Dinosaur Project data are owned jointly by the whole community, so this is no longer an issue.

Scientists lament the public's poor understanding and mistrust of science, and funding agencies want demonstration of 'broader impact' for research proposals. Public databases, particularly for engaging disciplines such as palaeontology and astronomy, may offer one solution.

Andrew A. Farke Raymond M. Alf Museum of Paleontology, Claremont, California 91711-2199, USA
e-mail: afarke@webb.org
Michael P. Taylor Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK
Mathew J. Wedel Western University of Health Sciences, Pomona, California 91766-1854, USA

Sharing: guidelines go one step forwards, two steps back

SIR — The newly revised *Publication Manual of the American Psychological Association* ranks among the most important books in the behavioural sciences and is one of Amazon.com's best-sellers in all book categories. Yet many research psychologists ignore its stipulations on the importance of data sharing.

The manual prescribes structure and style for scientific manuscripts, and deals with the ethics of publication. In this and earlier versions, authors are instructed not to withhold their raw data from other researchers who wish to verify the conclusions through reanalysis (citing standard 8.14 of the APA Ethics Code). It has been found, however, that 73% of psychologists publishing in high-impact APA journals failed to meet this obligation (J. M. Wicherts et al. *American Psychologist* **61**, 726–728; 2006).

Unfortunately, the revised APA guidelines in the new edition aggravate the situation. They stress that data should be shared by written agreement: the agreement must specify conditions relating to the proposed use of the data (for example, for verifying results or for secondary analysis), limits on dissemination of re-analysis results, and authorship expectations. Although imposing such conditions on other researchers seeking to use the data for their own ends is reasonable, it will hinder data-sharing for verification purposes.

Suppose that statistical flaws become apparent in a published paper. Under the new guidelines, these can't be investigated by an independent researcher unless the author agrees the terms for sharing the raw data. For instance, the author may demand to be a co-author on ensuing publications. It seems to us that the new APA guidelines will

impede, rather than advance, the critical assessment of the quality of data analyses in psychology research.

Jelte Wicherts, Marjan Bakker Faculty of Social and Behavioural Sciences, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands
e-mail: j.m.wicherts@uva.nl

UK red tape as sticky as the US version for would-be students

SIR — In a recent Editorial, you describe the arduous visa requirements applied to students or scientists seeking to enter the United States (*Nature* **461**, 12; 2009). I experienced similar challenges and more as an American applying for a student visa to enter the United Kingdom.

My permanent residency in Europe counted for nought. Bewildering tangle of directives? Check. Mysterious acronyms? Check. Proof of money to cover expenses? Check. Uncertain waiting times? Check. The UK visa system has requirements not (yet) dreamed of in the United States, such as a letter from the admitting university not only confirming that the student has been accepted but also detailing why. The exam scores or reference letters used in the admittance decision also have to be included in the application process.

It is not only the United States that needs to consider balancing the need for security with allowing students and scientists reasonable terms of entry to the country.

Colin Stoneking Heinrichstraße 25, 04317 Leipzig, Germany
e-mail: cjstoneking@gmail.com

Contributions to this page may be submitted to correspondence@nature.com. Please see the Guide to Authors at <http://go.nature.com/cMCHno>. Comments and debate are also welcomed at our blog Nautilus (<http://blogs.nature.com/nautilus>).

OPINION

India pushes for common responsibility

Rajendra K. Pachauri says that India wants to be a constructive partner in Copenhagen negotiations on climate change. The country is taking domestic action even though it cannot accept mandatory emissions limits.

India expects a strong agreement at December's United Nations climate conference in Copenhagen for several reasons. First, the country is very vulnerable to the effects of climate change, both those projected to occur within its own territory and those in neighbouring countries. Bangladesh, for instance, with a population of 160 million people is extremely vulnerable to sea level rise. This, along with the growing intensity and frequency of cyclones and other extreme events could result in large numbers of migrants fleeing to India. Equally serious are the problems associated with glaciers melting in the Hindu Kush region. Most of the rivers in northern India originate in these glaciers, and a decline in river flows because of reduced glacial mass would lead to water scarcity for India and its neighbours. Climate change is also likely to directly affect agricultural production, because there is growing evidence that some agricultural crops are seeing a decline in yields due to climate change, most notably wheat crops grown in North India (H. Pathak *et al.* *Field Crops Res.* 80, 223–234; 2003). Consequently, India has a vital stake in the stabilization of Earth's climate system.

Second, India is not solely interested in seeing global emissions adequately reduced as soon as possible. The country also has a direct interest in adaptation measures for coping with projected climate change, which is now inevitable because of the existing inertia in the climate system. India feels strongly that on the basis of historical responsibility and considerations of equity, developed countries should provide financial support for adaptation in developing countries. It is no coincidence that the subject of adaptation first emerged as a prominent issue in international climate negotiations held in New Delhi in 2002. Since then, the debate has gone further, with a number of developing countries joining in the demand for adequate financial support for this purpose from developed nations.

India has been one of the strongest proponents of the principle of 'common but differentiated responsibility' included in the United Nations Framework Convention on Climate Change, demanding action strictly on this basis. In India's view, therefore, the Copenhagen agreement should involve major cuts in greenhouse-gas emissions by the developed

countries, and should clearly articulate the support that the developed countries will provide for access to low-carbon technologies in the developing world. Over time, this will help developing countries implement measures to limit their own emissions.

Within the limited time available, India can perhaps play a constructive part in forging an agreement in Copenhagen. As a matter of principle, India will firmly dismiss demands from developed nations that their proposed emissions cuts should, in any way, be contingent on rapidly developing economies, such as India and China, committing themselves to emissions limits before 2020. But to demonstrate the country's seriousness towards shared action, India has a domestic programme — the National Action Plan on Climate Change (NAPCC) — that it could, at an appropriate stage of the negotiations, offer as part of a global package of commitments. The NAPCC was released by India's Prime Minister Manmohan Singh in June. It was inspired, in no small part, by the attention given to the 2007 Fourth Assessment Report of the Intergovernmental Panel on Climate Change.

The plan focuses on achieving a pattern of sustainable development while dealing comprehensively with the challenge of climate change. In Copenhagen, India should reject any imposition of measures for verification of goals achieved under the NAPCC, but might agree to annual international reporting. As a

democracy with a vibrant civil society and a free press, India can be comfortable with such a provision because it would not be revealing any information that wouldn't be publicly available anyway. Still, to achieve the ambitious goals of the NAPCC will require unprecedented institutional changes within India to overcome the bureaucratic inertia that often stymies government plans. In particular, a greater role would be desirable for public-private partnerships and closer involvement of think tanks and research organizations in providing policy advice.

Technological partnerships

Some might say that India's projected three- or fourfold growth of emissions by 2030 under business as usual (see graphic) is much too high an estimate, but given the high economic growth rates achieved by India in recent decades this is a realistic projection. Without the development of low-carbon technologies and the transfer of technology, developing countries will continue to emulate the emissions-intensive growth patterns seen in the developed world. But alongside well-defined intellectual property rights, technological innovation can be created through partnerships between developed and developing countries at much lower cost than if these were driven solely by the developed world. It is for this reason that US companies such as GE are carrying out substantial research and development activities in India.

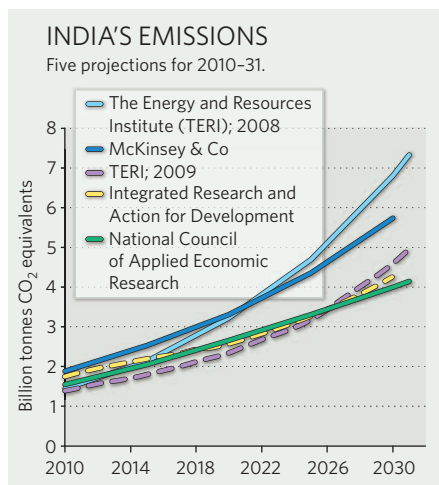
We need a more constructive spirit between developed and developing countries if we are to reach an effective agreement in Copenhagen. To achieve this, the leadership of developed nations must convince their public on the principles and scientific realities under which they will take proactive commitments. At the same time, major emerging economies such as India should not hesitate to put forward their own national action plans as part of global efforts, and which would help them achieve sustainable development. ■

Rajendra K. Pachauri is chairman of the Intergovernmental Panel on Climate Change, director-general of The Energy and Resources Institute and director of the Yale Climate and Energy Institute.

e-mail: pachauri@teri.res.in

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen.

Join the discussion at go.nature.com/hzQ2MD.



China expects leadership from rich nations

Greater emissions cuts by developed nations are the starting point for a successful climate deal at Copenhagen in December says **Jiahua Pan**.

To reach a successful climate agreement at the UN climate conference in Copenhagen, three fundamental elements (scientific evidence, political will and economic interest) and four practical elements (target setting for emissions reductions, adaptation, technology and financing) need to be addressed. Of these factors, negotiators will focus most attention, superficially, on mitigation targets. More fundamentally, however, they will focus on understanding economic impacts.

The Chinese perspective is that reaching a deal will depend largely on decisive mitigation action being taken by the developed nations. The developing nations will have every reason to follow suit if the rich nations demonstrate leadership and commit to more substantial cuts than they have offered so far.

There is already consensus on two of the fundamentals. The scientific conclusions outlined in the 2007 Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) have been largely accepted by the global community. This year, the heads of the G8 countries expressed support for further climate-change actions: committing to limit global warming to 2°C and proposing 50% global emissions cuts by 2050 from 1990 levels, with 80% cuts by industrialized nations (up from 50% a year earlier). It seems that there is no lack of political will. The only fundamental left is economic interest. Developed countries are concerned with immediate negative economic effects, whereas the developing countries are worried about their future well-being if they sign up to a legally binding, but unrealistic, target.

There are disagreements over all four practical elements, each linked to economic self-interest. Adaptation, technology and financing are important for a climate deal but these are, at most, incentives for developing-country participation. All nations require adaptation, but developed countries have adaptive capacity whereas the developing ones do not. Financial resources are thus required to help the poorer countries to adapt. Technology transfer also entails the transfer of financial resources, because technologies have a market value.

So far, little external funding has been forthcoming. Some development assistance will be provided by rich nations, but this is far less than what is needed. The uncertainty over financing is a major barrier to a global climate deal.

Setting mitigation targets is the most challenging practical element. According to the 2007 Bali roadmap, the global community must have a long-term shared vision, and industrialized nations should make deeper emissions cuts by 2020, whereas developing nations should take measurable, reportable and verifiable mitigation actions. The shared vision is often interpreted as a mitigation target for 2050.

But are the cuts proposed by rich nations sufficient? The G8 proposal, although promoting a 50% global cut for 2050, avoids a target for 2020. The ratio of emissions between the Annex I (industrialized) nations and rest of the world is roughly 50:50. If Annex I nations cut their emissions by 80%, developing nations will have to cut their emissions by some 20% in absolute terms from their current relatively low levels. In 2005, Annex I nations emitted almost five times the rate per capita of non-Annex nations. An 80% reduction would mean that by 2050, per capita emissions for Annex I nations would drop to the 2005 levels for developing nations.

Yet to meet the G8 proposal, the per capita emissions of developing nations would have to be 20% lower than this level. This would mean that per-capita emissions in the developing world would always be lower than those in the developed nations, in the past, now and in the future. As fossil fuels are cheaper than carbon-free energy sources, developing countries argue that a premature shift to low-carbon energy may slow development. Only a slow and limited switch to low-carbon sources has been seen in rich nations, despite their technological

and financial advantages. Even so, China has been investing in low-carbon energy sources at a higher rate than most rich countries.

A 2020 target for cutting emissions is even more important but elusive. The IPCC recommends 25–40% cuts by 2020 for Annex I nations, and for developing countries to lower their emission pathways 15–30% from business as usual. A 40% cut by 2020 means decarbonization at twice the rate needed to meet the G8's 2050 target. But so far, none of the Annex I parties has voluntarily committed to a 40% target. The European Union, as

front-runner, offered to boost its 20% cuts to 30%, conditional on other parties' actions. The United States hasn't ruled out a 2020 target, but it is expected to be well below 25%. Developing nations are willing to take mitigation action, but this is conditional on receiving technology and finance. Economics is preventing all nations from taking stronger action.

To achieve a solution, developed countries must show leadership in Copenhagen. They should promise cuts equal to, or deeper than, 40% for 2020. If the Annex I parties are unwilling or unable to do this, the rest of the world would be discouraged from taking serious action. A more likely outcome in Copenhagen would be a statement that the world intends to limit global warming to 2°C by 2050. Emission reductions and mitigation actions for individual parties will have to be specified later.

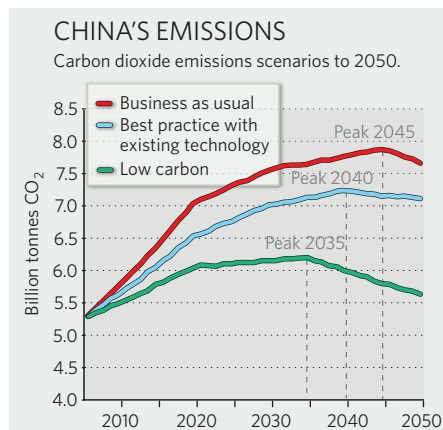
Even so, the developing countries should prepare nationally appropriate mitigation plans for low-carbon development. These could include reducing carbon emissions per unit of gross domestic product and making tougher renewable-energy targets, both of which are compatible with development. It is in this regard that China and India as large developing economies can push forward the process. The road to and beyond Copenhagen will be tough, but there is no alternative. ■

Jiahua Pan, Institute for Urban & Environmental Studies, Chinese Academy of Social Sciences, 10 Xianxiao Hutong, Dongcheng District, Beijing, 100005, China

e-mail: jiahuapan@163.com

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen.

Join the discussion at go.nature.com/hzQ2MD.



Copenhagen needs a strong lead negotiator

Reaching an international climate agreement requires someone with exceptional skill, knowledge and diplomacy, says Kyoto chair **Raúl Estrada-Oyuela**.

As the professional diplomat who presided over the Kyoto Protocol negotiations, it is clear to me how vital it is to have a good leader to steer negotiations at the Copenhagen conference on climate this December. Time is short, and matters are very complex. Although it may prove impossible to agree on quantified commitments at the meeting itself, a strong effort should be made for a deal that at least settles the main political objectives, with the aim of finalizing the agreements at subsidiary meetings in June 2010.

The United Nations Framework Convention on Climate Change (UNFCCC) currently involves 192 governments, making negotiation a major exercise in diplomacy. To come to an agreement, the process needs a well established leader who is fair, forceful, committed and well informed on the subject under debate and on the aspirations and bottom lines of all parties.

Often, the host country can take on this role. But Denmark, the host of December's 15th Conference of the Parties to the UNFCCC, has pushed the objectives of the European Union so aggressively that a leadership role from that country is likely to generate a negative reaction from some parties, such as India. The difficulty of Denmark's position in leading the talks was emphasized on 12 October when the country's chief climate negotiator Thomas Becker left his post in the wake of an expenses scandal. Plus, the nation's intended leader of the conference — Anders Fogh Rasmussen — resigned as prime minister earlier this year to become NATO's secretary-general. The current Danish minister of energy and climate — Connie Hedegaard — will therefore officially preside over the talks, but it is not clear to me that she has the necessary experience to truly lead the negotiations.

A whole solution

The best option, I think, would be to create a Committee of the Whole that would combine negotiations from the UNFCCC and the Kyoto Protocol, which are currently run by separate committees with separate chairs. It is still possible to do this, and it will remain possible until the conference starts. The last preparatory meeting, in Barcelona on 2–6 November, will be an opportunity to consider this point. If such a committee is formed, its elected chair would naturally lead the negotiations at the Copenhagen summit. If this doesn't happen,



Diplomats must wrangle negotiations deep into the night, as Raúl Estrada-Oyuela (right) does here during the Kyoto Protocol discussions in 1997, with Michael Zammit Cutajar (middle) and Richard Kinley (left).

the chance of making a good deal will be lessened.

My own role in climate negotiations began with the Second World Climate Conference, held in Geneva in 1990, where I was Argentina's representative and the de facto speaker for the Group of Latin America and Caribbean countries. A graduate in law, I had joined the foreign service in 1966 and participated in a number of multilateral negotiations, including on environmental issues. Diplomats at the 1990 conference were, surprisingly, not allowed to attend the science segment devoted to the first assessment report of the Intergovernmental Panel on Climate Change (IPCC). Before the conference, from my desk in Buenos Aires I had to send junior colleagues to attend meetings they barely understood in Geneva to get a handle on the IPCC's activities. Eventually I was lucky enough to recruit a good scientific team, led by Osvaldo Canziani, to educate myself and my colleagues.

I learnt many good lessons from other diplomats about the skills required to negotiate a good agreement. Jean-Maurice Ripert, for example, was a distinguished French economist

who was ambassador to the UN and chaired the Intergovernmental Negotiating Committee in 1991. This was the body tasked with creating a legal multilateral instrument on climate change (what came to be called the UNFCCC). I was vice-chairman of that committee at the time, and witnessed Ripert's skill first-hand. He was an optimist in the most adverse of circumstances. He was well aware of the need to have the United States and Japan on board, yet also had authority among developing countries, because he was in charge of promoting their participation on the IPCC. He consulted privately with most delegations on every issue to understand their thinking.

Also a master of this trade is ambassador Tommy Koh — former dean of the Faculty of Law at the National University of Singapore, a UN representative and eventual chair of the preparatory committee for the Earth Summit in Rio de Janeiro in 1992 (and chair of the Committee of the Whole in that conference). During the Earth Summit negotiations, he commanded daily results from the leaders of sub-groups he had created. When he felt that

K. KASAHARA/AP

an area was failing to progress, he took the matter on himself, making emotional appeals in the plenary if necessary and even jokingly comparing his appearance to Mickey Mouse to soften the debate. The Rio Declaration on Environment and Development is, in large part, a product of his personal drafting.

Clever tactics

The Intergovernmental Negotiating Committee met for the first time in February 1991 just outside Washington DC. From the start, as today, the United States and India made progress difficult: at one point, the US representative, Bob Reinstein, refused to participate in a crucial meeting. After insisting in many ways without success, I obtained a suite for an impromptu luncheon and invited all the relevant delegates, including Reinstein. Using a trick learned from Koh, I appealed to their good manners and the rules of polite diplomacy to point out that they could not decline the chairman's invitation. We had lunch, we forged an agreement and, fortunately, I had a credit card to cover the bill.

Final success in drafting the UNFCCC was reached only because Ripert worked full time to host a multitude of consultations during and between the official sessions. Once each paragraph of the convention was completed, we still needed to obtain consensus on the package. In delegates' jargon, 'consensus' means that everybody can live with the text even if not fully satisfied by it; it is reached not by a vote, but by the lead negotiator taking the responsibility to declare that a consensus has been reached. OPEC (Organization of the Petroleum Exporting Countries) members vehemently held back their consent, but Ripert managed to moderate their resistance to a point at which he felt justified in declaring a consensus. That took guts.

The UNFCCC then went to the Rio Earth Summit in 1992, where it was signed into being. Further work was needed to create a set of rules by which it would be enacted. This again required a series of meetings by the Intergovernmental Negotiating Committee, for which I was elected chair.

The main product of the first meeting, held in 1995, was the Berlin Mandate, which established the basis for the negotiation of the Kyoto Protocol. This complicated document was created under the leadership of ambassador Bo Kjellén and Angela Merkel, then Germany's environment minister. Merkel is a superb politician. Having grown up in eastern Germany, she was versed in the uselessness of inflexibility and devoted to constructive compromise. Merkel worked a whole night as president of the conference,

shuttling from one room to another, to work out the final text at around 6 a.m. of the last day. Like Ripert before her, she had obtained a strong enough position to declare the mandate adopted by consensus despite protest from the OPEC members.

We worked on the text for the protocol from August 1995 until December 1997, with the final stage in Kyoto. Although Japan held the presidency for this conference, it was unable to field a player similar to Merkel because of internal squabbles between the ministries of foreign affairs, international trade and industry, and environment (this continues to be the case in Japan). Instead, I took the lead.

Having studied Koh, Ripert and Merkel, I used diverse instruments to steer the discussions, always consulting with all sectors. At the request of the United States, we slowed the process until President Bill Clinton was re-elected. To placate the constant claims of the OPEC countries, I asked my Iranian friend Mohammad Reza Salamat, now a programme officer in the UN Secretariat, to find a way to placate the oil producers (which he did by drafting two paragraphs of the protocol). The United States favoured a 'cap and trade' approach from the beginning, allowing the market to drive emissions cuts, whereas the European Union was initially more inclined to adopt 'policies and measures', relying on rules and regulations to reduce emissions. We created an elaborate mix, with 'cap and trade' dominating, to satisfy all.

No regrets

The political decisions made for the Kyoto Protocol had their shortcomings. We opted to put targets on a range of gases, knowing that we had very different degrees of certainty on the estimation of emissions for each of them.

Selecting 1990 as the base year for emissions comparisons was arbitrary, if politically convenient, as was selecting a 100-year horizon. The target of a 5% emissions reduction was too modest, but it was the only one politically possible. Despite this, the Kyoto Protocol has had an undeniably positive impact in international policy: climate change is now at the centre of the international scene, and I have no regrets about the negotiation.

Some have suggested that large deals always result in unsatisfactory compromise, and that smaller agreements are better alternatives. The administration of former US President George W. Bush, for one, prompted meetings among some 15–20 governments to discuss their

own climate initiatives. These talks continue, but with no results. Differences between large world economies remain the same even in small meetings. Plus, only large meetings can properly capture the needs of the developing world.

Who will be Copenhagen's Merkel or Ripert? I hope this will be established before December.

The main task for Copenhagen is shared by two groups. The working group on long-term cooperative action must propose a "comprehensive process to enable the full, effective and sustained implementation of the Convention", with the intention of including commitments from the United States and the 'mega' developing countries — China, India, Brazil, Mexico and North

Korea — that do not have quantified commitments in the Kyoto Protocol. In addition, the working group on further commitments for 'Annex I' countries under the Kyoto Protocol must propose targets for beyond 2012 for developed nations who had caps for their emissions in the 2008–12 commitment period.

This March, ambassador Michael Zammit Cutajar of Malta, former executive-secretary of the UNFCCC, took the chair of the long-term co-operative action group. Cutajar has the necessary capabilities and the knowledge to lead this group. His nation is a member of the European Union but it is in many aspects a developing country, giving him respect from many quarters. In June, he produced an excellent draft negotiating text of fewer than 60 pages based on proposals submitted by various governments. The document has not met with much success, however: government representatives have added paragraphs that push the text up to 200 pages.

At the same time, ambassador John Ashe, of Antigua and Barbuda, took the chair of the further commitments working group. He, too, has produced a draft negotiation document and has considerable experience as a chairman, but he too has been confronted with a lack of cooperation by the parties.

Hedegaard has a difficult task at Copenhagen. It might be advisable not to end the conference in December at all, but rather, as we did with the 6th conference at The Hague, reconvene it six months later. Such a delay is not the best option, but may be the only way to reach a meaningful agreement.

Raúl Estrada-Oyuela spent 42 years as a diplomat for Argentina. He has now retired in Buenos Aires.

e-mail: eoy@estrada-oyuela.com.ar

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen. Comment on this article at go.nature.com/hzQ2MD

"The Kyoto Protocol has had an undeniably positive impact in international policy."



BOOKS & ARTS

Conveying the campaign message

The arts and advertising can galvanise public and political will in tackling global warming. But shared concern for human health is a better motivator than polar bears, finds **Sanjay Khanna**.



This August at Beijing's Temple of Earth, an installation of 100 ice sculptures of children melted in the heat. It represented the billion lives that will be lost in Asia because of water shortages caused by climate change. The artwork — commissioned by Greenpeace to launch the TckTckTck campaign for 'bold climate action' — is one of many cultural events aiming to sway political negotiators in the lead-up to the United Nations Climate Change Conference (COP15) in Copenhagen in December.

"Arts and cultural experiences may be the most effective and powerful ways to communicate the impacts of climate change to a significant portion of the population," says atmospheric scientist David Battisti of the University of Washington in Seattle.

Indeed, many climate campaigns are using arts approaches within a raft of other public-awareness strategies. Artists, writers and musicians are also commenting independently on the potentially grave consequences if governments do not agree to curb greenhouse-gas emissions.

But for climate campaigners to be influential, amid the realpolitik of governments and the din of society, their point must reach and be adopted by a wide audience. Spreading a culture-based message is one of many tools used by the public-relations and advertising industry, which calls on decades of psychological and sociological research to understand how people's attitudes form and change. Historically, such efforts have been directed towards building political following or promoting governmental and business interests and products. Today, the same techniques of persuasion could hold the key to increasing climate-change awareness and ameliorating the cognitive impact of decades of advertising.

Artists, skilled in conveying ideas through the senses, can have an influential role in shaping public opinion about climate change. When engaging with the arts, "people expect to be in the realm of their emotions and of mystery and metaphor, and this is fertile ground for planting seeds of change", notes mezzo-soprano and theatre producer Miranda Loud, who founded the multimedia arts group Rialto Arts near Boston, Massachusetts. Loud's award-winning

production *Buccaneers of Buzz* highlights the devastating effects of climate change on bees.

The cross-fertilization of ideas between artists and climate scientists is being fostered by international projects such as Cape Farewell. Set up in 2001 by the artist David Buckland as a non-profit organization and partly funded by universities, the project runs interdisciplinary expeditions to the Arctic and South America. Participants have included novelists Vikram Seth and Ian McEwan, as well as musicians Leslie Feist and Laurie Anderson. The fruits of such efforts might not always be immediate, but can reach many: for example, McEwan's next book, due out next year, will centre on climate change.



Melting ice sculpture highlights impact on humans.

Iconic photographs of polar bears clinging to ice floes have elicited widespread public sympathy, yet few people have been moved enough to alter their behaviour to become more 'green'. Environmental photographer Matthias Klum thinks that artists should do more to influence the public: "Art is a tool that isn't used enough to effect change," he maintains. During COP15, Klum will unveil *The Testament of Tebaran*, a photographic exhibition that illustrates the effect of deforestation on carbon dioxide levels.

When it comes to persuasion, film is an effective format that can be widely distributed. Al Gore's *An Inconvenient Truth* (2006) reached an audience of millions and helped earn him a Nobel peace prize. Several

documentaries released in the past year reflect on what we stand to lose through inaction on climate change. *Home* (2009), by noted aerial photographer Yann Arthus-Bertrand, evokes sadness at the gradual destruction of familiar landscapes and lifestyles. Director Franny Armstrong's *Age of Stupid* (2009) examines the despondency felt by people in 2055 when they look back and wonder why their predecessors didn't protect the planet from climate change when it was still possible.

This message of urgency is harnessed by the TckTckTck campaign, backed by organizations including Amnesty International, Oxfam, the Union of Concerned Scientists and the Global Humanitarian Forum led by former UN secretary-general Kofi Annan. Time literally ticks away on the campaign's online clock as it counts down to the start of COP15. The campaign's song, released this month, is a version of the 1987 hit 'Beds are Burning' by Australian band Midnight Oil, who have rewritten and recorded it with more than 60 artists and celebrities including Duran Duran, Lily Allen, Bob Geldof and Youssou N'Dour. Midnight Oil's front man Peter Garrett is now Australia's environment minister.

This cultural clamour for radical action could still get shouted down by those who are lobbying equally vociferously — and with infinitely greater financial backing — to keep the status quo. James Hoggan, public-relations strategist and author of *Climate Cover-Up* (Greystone, 2009), claims that well-orchestrated, big-budget efforts have stoked unfounded controversy: "It's beyond David and Goliath," he says.

Spanning diversity

To find a global consensus, conversations on climate need to span diverse groups and surmount linguistic and cultural barriers. Campaigns that use numbers are one option: "Slogans can mean something different in Johannesburg and Delhi and Vancouver, but numbers remain reassuringly the same," says Bill McKibben, author of *The End of Nature* (Random House, 1989). He leads the campaign 350.org, which advocates decreasing atmospheric CO₂ to 350 parts per million. On 24 October, his group will seek media attention by staging visual stunts that represent the number 350 through group performances, including choreographed mountain climbing and skydiving. "Physics and chemistry are poor

negotiators — they don't haggle or compromise," reflects McKibben. Tied to her film *The Age of Stupid*, Armstrong also leads a number-based campaign: '10:10' calls for a reduction in UK carbon emissions by 10% in 2010.

Human rights is an area of universal concern that can be effective in building consensus, as exemplified by Beijing's ice sculpture. Another motivator is the ethical pressure that lies behind the language of green marketing. Although sustainably designed products promote consumption through their desirability, they offer an easy route for individual action. As a prelude to COP15, Copenhagen last month hosted a design week to showcase sustainable Scandinavian flair, from fair-trade products and energy-saving devices to city parks and cycleways.

Audience fragmentation, a long-standing legacy of marketing, must also be overcome by communicators. Online multimedia projects can reach diverse audiences that are used to being selectively targeted by the media. One such project is *100 Places to Remember Before They Disappear*, organized by Stine Norden and Søren Rud of Danish communications firm Co+Life, which combines a website with short television programmes, educational



Women queue for water at a well near Khudiala in Rajasthan, India, in Yann Arthus-Bertrand's film *Home*.

materials and a travelling photographic exhibition to raise awareness and present climate solutions. It highlights characterful locations that have been identified by the Intergovernmental Panel on Climate Change as being at risk of loss in the next 30–60 years — including The Battery in New York City; Rotterdam in the Netherlands; Mount Kilimanjaro in Tanzania; and Perth, Australia.

So far, the cultural climate debate seems to have been dominated by palpable environmental concerns such as melting glaciers — perhaps because of the dramatic visual evidence they

provide. But the loss of place and habitat affects people's well-being and resilience, bringing with it the severe threat of drought and disease. Promoting such concern over global health would put a human face on climate change — this may be the impetus we need to form a worldwide consensus for action. ■

Sanjay Khanna is a climate-change writer and journalist based in Vancouver, Canada.
e-mail: sanjaykhanna1@gmail.com

See Editorial, page 1027, and online at www.nature.com/roadtocopenhagen



Q&A: The science of persuasion

Psychologist **Robert Gifford** is co-author of a recent American Psychological Association report that examined the interface between psychology and climate change. He explains what makes people receptive and how to get messages about climate science across effectively.

Why do climate scientists who speak to the public face a paradox?

Scientists understand that climate models inherently have uncertainty. But as soon as the public experiences some uncertainty in the message, people begin to dither and to say, "Well, I guess it's not so urgent — the experts aren't so sure of themselves".

How can climate campaigners make themselves heard?

Messages have to be carefully targeted to different demographic groups, based on their self-interest. For the older generation, it's about their children and grandchildren; and for the younger generation, it's about them and their generation.

How useful are phrases such 'climate justice'?

For people with social-justice concerns, such a phrase will work. It resonates with members of Greenpeace and of left-leaning political

parties; people more on the right side of the political spectrum will say, "You take care of your problems, we'll take care of ours".

What five elements make up an effective message?

First, it has to have some urgency. Second, it has to have as much certainty as can be mustered with integrity. Third, there can't be just one message: there must be messages targeted to different groups. Fourth, messages should be framed in positive terms. Evidence from a recent thesis I co-supervised shows that people are less willing to change their behaviour if you tell them they have to make sacrifices. If you tell them they can be in the vanguard, be a hero, be the one that helps — that works. Fifth, you have to give people the sense that their vote counts and that their effort won't be in vain.

If some commentators are right and we are past the point of no return, how could

we mobilize people into action?

The same five elements would apply, except that the messages would emphasize adaptation and mitigation. We would need to tell people they can lessen the impact and be someone who helps our own and other species — but with a tilt towards communicating that there will be huge impacts.

How do we communicate difficulties while making clear there are myriad options to be constructive?

We can learn from the success of other campaigns, such as seat-belt, anti-alcohol or anti-smoking campaigns. The fear-appeal approach can be overdone to the point where it has a counteractive effect. Things should be presented more in terms of asking how each person is going to help us face the challenge. Saying "It's all over" makes people give up. ■

Interview by **Sanjay Khanna**.

See report at <http://tinyurl.com/yfaobrg>

Climate books with clout

David Reay examines the evolution of books about global warming and highlights those that have had most influence on public perceptions.



In the past two decades, a plethora of global-warming books has catered for increasingly sophisticated and specialist demand. Science, policy and public opinion evolve daily, so any new book on climate change risks being outdated before it hits the shelves. Nevertheless, some accounts have been highly influential.

The first popular climate-change books were generalists. Released into a world with few competitors, Bill McKibben's *The End of Nature* (Random House, 1989) and John Houghton's excellent *Global Warming: The Complete Briefing* (Lion Publishing, 1994) flourished as an increasing number of readers looked for a digestible take on the dense tomes of the newly formed Intergovernmental Panel on Climate Change (IPCC), of which Houghton was co-chairman. These books used the blunt declaration that anthropogenic global warming is real to pull in a lay audience whose awareness of climate change was only just beginning to build.

Then, against the backdrop of the Kyoto Protocol negotiations in the late 1990s, and with the first nudges of selection pressure from readers, climate-change books began to speculate. Expanding niche audiences demanded texts that were tailored to their own demographics, ethics and politics. As the new millennium dawned, there was the literary equivalent of a Cambrian explosion.

Vivid palaeoclimate narratives such as *The Two-Mile Time Machine* by Richard Alley (Princeton University Press, 2000) and impact travelogues like Mark Lynas's superb *High Tide* (Flamingo, 2004) shouldered their way into the territory of popular science. The genus of 'individual action' rapidly expanded, its variants ranging from Mayer Hillman's puritanical *How We Can Save the Planet* (Penguin, 2004) to the gentle hectoring of titles including Leo Hickman's *A Good Life* (Eden Project Books, 2005) and my own *Climate Change Begins at Home* (Palgrave Macmillan, 2005).

Meanwhile, manifestos illuminated by policy proliferated, such as *Contraction and Convergence* by Aubrey Meyer (Green Books, 2000) and

George Monbiot's incisive polemic *Heat* (Allen Lane, 2006). Along with them came contrarian lambasts including *Hot Talk, Cold Science* by Fred Singer (Independent Institute, 2001) and wide-mouthed critiques such as Bjørn Lomborg's *The Skeptical Environmentalist* (Cambridge University Press, 2001). Each found its own oasis of readers. With new editions, columns and serializations, they could draw in new readers or occupy another's territory. None could yet lay claim to a true mass audience. Then, in 2006, along came *An Inconvenient Truth*.

Al Gore's book (Bloomsbury), written to accompany his blockbusting film, quickly topped *The New York Times* non-fiction best-seller list. Nothing has matched it in terms of global popularity, although big hitters such as *The Economics of Climate Change* by Nicholas

cutting emissions than there is time to act. On the projected impacts of warming, Lynas's *Six Degrees* (Fourth Estate, 2007) is hard to beat. But as for books on adaptation to a warmer planet, the literary landscape remains sparse. And with increasing pressure to reduce emissions and improve impact projections, more topics are still to come under the literary spotlight.

When I canvassed my carbon-management masters students on their favourite climate books, responses sampled most of the spreading family tree. Plaudits came in for Nigel Lawson's sceptical *An Appeal to Reason* (Overlook Press, 2008) alongside those for Stern's *A Blueprint For A Safer Planet* (Bodley Head, 2009); James Garvey's *Ethics of Climate Change* (Continuum, 2008); and Chris Goodall's *Ten Technologies to Save the Planet* (Green Profile, 2008). The common thread is that they are all recently published, well-written books that increase understanding and provoke debate — whether or not the thesis tallied with the reader's.

If books have changed public attitudes to climate change in the past two decades — and I believe that they have — then it is those that have challenged our preconceptions and taught us something new that deserve the credit. One of the triumphs of Gore's *An Inconvenient Truth* and Lomborg's *The Skeptical Environmentalist* was the intense debate that each engendered. The very process of claim slugging it out with counterclaim on Internet message boards and newspaper correspondence pages increased awareness and understanding of the issues being discussed.

In the future, the public might opt for dynamic electronic-book titles that can hit screens with up-to-date information and respond to readers' queries. The carbon footprint of 'dead-tree' editions counts against them too, with the extra burden of emissions from production and distribution giving their e-competitors another boost.

In this market, there is little room for generalists without a big name. Newcomers must rely more on the publicity fillip of controversy, or on translating large online followings into book buyers — as with David McKay's surprise hit *Sustainable Energy* (UIT, 2008) and Greg Craven's *What's the Worst that Could Happen?* (Perigee, 2009), both YouTube-powered examples.

Another *Inconvenient Truth* may yet spring from the tangle of climate book publishing. Still, there is only one author who has sufficient fame, influence and writing talent to do it on a world-changing scale. His name is Barack Obama. ■ **David Reay** is a senior lecturer in the School of GeoSciences, University of Edinburgh, UK, and author of *Your Planet Needs You!*. e-mail: david.reay@ed.ac.uk



Stern (Cambridge University Press, 2007) and updated instalments of the *IPCC Assessment Report* (Cambridge University Press, 2007) have arguably had a greater impact on global policy.

The evolution of climate books shows no sign of stopping. Recent years have seen the success of 'apocalypse-soon' titles, such as James Lovelock's *The Revenge of Gaia* (Allen Lane, 2006) and Fred Pearce's *The Last Generation* (Eden Project Books, 2006). Today, no big policy-maker's CV is complete without at least one climate-change book — Gabrielle Walker and David King's cracklingly good *The Hot Topic* (Bloomsbury, 2008) being my pick of the bunch.

Gore has penned a version of *An Inconvenient Truth* for the young adult market, and the current best-seller lists of global-warming titles include those aimed at audiences from churchgoers and corporations to doctors and climbers. There are now far more books to read on



A sustainable satellite development near Shanghai.

Q&A: Architect of a sustainable future

German architect **Albert Speer Jr** is a pioneer of sustainable building and city planning whose firm has designed ecological communities from Cologne in Germany to Shanghai in China. With the publication of a new book setting out his philosophy, he explains why we should take a more holistic approach to urban development.

What does sustainability mean to you?

That we don't use more resources today than we can rebuild for the next generation. We still, everywhere, need too much energy and too much of everything else — mobility, land use and so on. Western society thinks in sectors, and not enough in circles. We must be holistic and interdisciplinary.

Are you optimistic about the power of architects to shape sustainable cities?

Architects and planners control roughly 5% of all the decisions needed for governing a city; the rest are political or based on different economic interests. You need an interdisciplinary team, which is missing in many governmental bodies.

The world is taking too long to make decisions. It's taken nearly two decades since the conference in Rio de Janeiro to get sustainability into the minds of politicians and governments.

China is building lots of new cities — what is their approach?

From my experience, I feel that the Chinese culture is better organized to manage problems than others. We are designing new eco-cities in China, where we design the infrastructure and urban planning. The main aim is to protect the environment and reduce the amount of farmland converted into cities. Chinese thinking is direct and practical, and they make decisions fast. My impression is that not everything is going in the right direction, but they are learning rapidly and want to build things in another way than in the past.

Are there differences between working in developed and developing countries?

In general, the problems are the same, but the solutions are very different. In developing countries, there is a lot of urbanization because cities are growing. In Europe and

other countries, the population is declining, cities are not growing any more, so the goals have to be more concentrated on existing city parts, rather than on the development of cities and new cities.

Does every place need its own approach?

Concerns about the environment, climate, land consumption, transportation and civic participation are the same everywhere. Architecture will be similar around the world



Architect Albert Speer pioneered 'green' planning.

in the future because people need similar spaces. But every city needs to think more about its own character and opportunities, as well as its landscape, culture, history and future. Integrated urban development will become more important, because cities are all competing against each other.

Is this true of 'megacities' such as Tehran, Manila and Cairo?

If you look at these cities in more detail, they're not one place with 10 million or 50 million people, they're 15 or 20 cities with 800,000 to 2 million people each. It's unrealistic to stop their development: we have to see them as polycentric so that they can function better.

For example, I'm involved in a project that started 25 years ago to restructure a city

of 3 million people within Cairo. It's called the City of 6th October and it's just behind the pyramids. Many things have been built, but public transport is missing, there's no green belt, and urban planning has been very scattered. Our basic plan has been adopted this year, and we're now in the phase of coordinating the development.

Do cities require social engineering too?

Yes. Urban development and sustainable cities have a lot to do with the people living in them. In Cologne last year, we came up with a master plan for the inner city of 130,000 people, and civic participation was a large part of our work. At first the interest was minor, but after a year our sessions were always full. Participation can't be stimulated by one public appearance and a lecture, it has to continue over several years.

What have been the biggest changes during your career?

Holistic thinking is more accepted. When I began trying to plan ecological cities 30 years ago, the word 'ecological' hardly existed. Because of technology and globalization, it's now easier to compare your own situation with what is happening elsewhere.

What's your favourite city?

In Europe, it's Barcelona. All the things we are talking about — transport, green space and so on — are integrated into the urban tissue. In Asia it's Singapore, which is a very dense city, but also green and very liveable. ■ Interview by **John Whitfield**, a writer based in London.

Albert Speer & Partner: A Manifesto for Sustainable Cities

by Jeremy Gaines and Stefan Jäger
Prestel: 2009. 240 pp. £25

NEWS & VIEWS

CHEMICAL PHYSICS

Molecular conformations fielded

Albert Stolow

Studies of molecular dynamics can be foiled by the presence of stereoisomers — molecules that have the same bond sequence arranged in different geometries. This problem has now been deflected.

In the ancient Indian tale of six blind men encountering an elephant, each man approaches the creature from a different direction. One finds the trunk, another a leg, the third the tail, and so on, whereupon they naturally disagree with each other as to the true appearance of the elephant (Fig. 1). And so it is with molecules — molecular properties and dynamics seem to depend on the direction from which you view them. To make matters worse, some molecules can exist as different stereoisomers — molecules that have identical atoms and sequences of bonds, but which have bonds that point in different directions. If elephants came in two versions, one in which the legs point towards the ground and another in which they point towards the sky, these two versions would be stereoisomers. Such a situation would clearly further complicate the deliberations of the six blind men.

Similarly, the presence of stereoisomers within a sample of molecules can lead to confusion in studies of molecular dynamics, because any measurements taken are generally averages of the properties of all of the stereoisomers. Because the dynamics of stereoisomers can vary greatly (as would the relative locomotive abilities of our elephantine stereoisomers), the average result may not pertain to any molecule at all. Reporting in *Angewandte Chemie*, Filsinger *et al.*¹ describe an approach for spatially separating stereoisomers of neutral molecules using electric fields. This avoids the detrimental averaging of results, opening up a way to study molecular dynamics that could fully benefit from emerging ultrafast measurement techniques.

As is often the case, the most powerful techniques rely on a simple premise. Filsinger and colleagues' method relies on a phenomenon discovered by Otto Stern and Walter Gerlach in 1922. In a classic experiment², Stern and Gerlach demonstrated that a beam of silver atoms was deflected in a magnetic field, a finding that hastened the birth of quantum mechanics. In a subsequent paper³, Stern suggested that electric fields could be used to spatially separate quantum states of neutral molecules. His proposal was based on an electric property of molecules known as the dipole moment. The basic idea



Figure 1 | The blind men and the elephant. Each man experiences something different depending on his angle of approach, in the same way that measurements of molecular properties seem to depend on the direction from which the molecules are viewed.

is that molecules 'feel' electric fields differently, depending on the value of their dipole moment. The stronger the electric interaction of the dipole moment with the field, the larger the possible deflection of a beam of molecules from its otherwise straight-line path.

The fact that some molecular bonds in stereoisomers point in different directions can lead to different dipole moments for each isomer. Filsinger *et al.*¹ recognized that this is the key to spatially separating stereoisomers using Stern's electric-deflection technique. For demonstration purposes, Filsinger *et al.* studied 3-aminophenol, which comes in only two stereoisomers (designated as *cis*- and *trans*-isomers) that differ in the orientation of an oxygen–hydrogen bond (Fig. 2, overleaf). Each stereoisomer has a different dipole moment. The authors show convincingly that

the *cis*- and *trans*-isomers in a molecular beam are deflected by different amounts in electric fields, and that this effect can be used to physically separate the isomers. The simplicity of this approach speaks volumes.

Until now, stereoisomers in mixtures could be selectively detected using highly resolved spectroscopic methods, but this did not allow stereoisomers to be spatially separated for practical use. Filsinger and colleagues' technique is therefore immensely valuable, although it does have limitations: it requires the stereoisomers to have sufficiently different dipole moments, and works best when the molecules are extremely cold, which can be difficult to achieve. Another method for separating stereoisomers in molecular beams using electric fields has previously been reported⁴, but the fields concerned were relatively complex. The advantage of the present deflection technique¹ is that it is simple enough to be readily used by researchers working in a range of other fields — and therein lies its importance.

There is a worldwide effort to develop new approaches for studying ultrafast molecular dynamics, with good reason. In the twentieth century, much of the progress in understanding the molecular underpinnings of our world focused on the basic structures of molecules, and the relationship of structure to function. Molecules, however, are seldom static. They often undergo dynamic processes, such as chemical reactions, and so a purely structural view is unlikely to be sufficient. The next step must therefore be to understand the dynamics of molecular processes.

Unfortunately, dynamics–function relationships do not yet exist. Many new ultrafast techniques are thus being proposed by which to study the 'structural dynamics' of molecular processes — the changes in position of atoms in molecules measured as a function of time. Methods such as time-resolved X-ray⁵ or electron diffraction^{6,7} and X-ray absorption⁸ can directly map out atomic motions. Other time-resolved techniques are able to observe both atomic motions and electronic rearrangements, and even their coupling as bonds rearrange in concert with atomic motions — for example, photoelectron spectroscopy^{9,10} (which studies

the distribution of electrons emitted from molecules when they are irradiated with ultraviolet or soft X-ray light) and high harmonic generation^{11,12} (which studies the burst of ultraviolet or soft X-ray light emitted from molecules when they are irradiated with intense, near-infrared laser pulses).

All of these technical developments, however, suffer from the same problem — averaging over stereoisomers and molecular orientations,

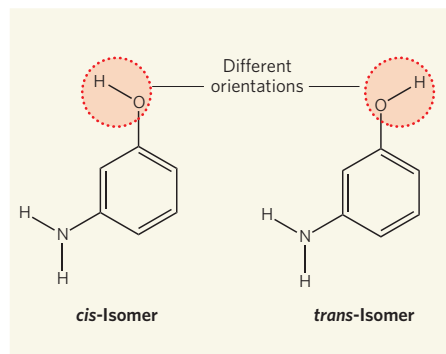


Figure 2 | The same, but different. 3-Aminophenol exists as two stereoisomers — *cis*- and *trans*-isomers — which differ only in the orientation of the oxygen–hydrogen bond (highlighted). The isomers are difficult to separate, but Filsinger *et al.*¹ have done so by deflecting a molecular beam of 3-aminophenol in an electric field.

leading to a blurring and potential loss of information. An analogy can be made with X-ray diffraction in crystals: the random orientation of crystallites in powder diffraction obscures the detailed information that is obtainable from single-crystal diffraction. Unless this averaging problem is seriously addressed, the new ultrafast techniques for studying molecular dynamics may not live up to their promise. Filsinger and colleagues' spatial selection of stereoisomers is therefore a step towards the full implementation of techniques that aim to illuminate the molecular processes that transform the worlds within and around us. ■

Albert Stolow is at the Steacie Institute for Molecular Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada.

e-mail: albert.stolow@nrc.ca

1. Filsinger, F. *et al.* *Angew. Chem. Int. Edn* **48**, 6900–6902 (2009).
2. Gerlach, W. & Stern, O. *Z. Phys.* **9**, 353–355 (1922).
3. Stern, O. *Z. Phys. A* **39**, 751–763 (1926).
4. Filsinger, F., Erlekam, U., von Helden, G., Küpper, J. & Meijer, G. *Phys. Rev. Lett.* **100**, 133003 (2008).
5. Chapman, H. N. *et al.* *Nature Phys.* **2**, 839–843 (2006).
6. Ihee, H. *et al.* *Science* **291**, 458–462 (2001).
7. Siwick, B. J., Dwyer, J. R., Jordan, R. E. & Miller, R. J. D. *Science* **302**, 1382–1385 (2003).
8. Bressler, Ch. *et al.* *Science* **323**, 489–492 (2009).
9. Geßner, O. *et al.* *Science* **311**, 219–222 (2006).
10. Bisgaard, C. Z. *et al.* *Science* **323**, 1464–1468 (2009).
11. Itatani, J. *et al.* *Nature* **432**, 867–871 (2004).
12. Li, W. *et al.* *Science* **322**, 1207–1211 (2008).

SOLAR SYSTEM

Saturn's colossal ring

Matthew S. Tiscareno and Matthew M. Hedman

A hitherto undetected disk of debris around Saturn is the largest ever found to be orbiting a planet. This ring may hold the key to one of the most enigmatic landscapes in the Solar System.

On page 1098 of this issue, Verbiscer and colleagues¹ report the discovery of an enormous ring around Saturn. The authors found this most tenuous of Saturn's known rings, which covers some 10,000 times as much area as the planet's photogenic main rings, by using the Spitzer Space Telescope to detect the ring's faint glow in the thermal infrared region of the electromagnetic spectrum. The ring is composed of dust, probably derived primarily from Saturn's distant moon Phoebe. Its discovery lends support to an earlier theory that dusty material from Phoebe is responsible for colouring the two-toned moon Iapetus.

This 'Phoebe ring' of Saturn (Fig. 1) is similar to previously known tenuous rings in the outer Solar System, such as Jupiter's gossamer rings or Saturn's E ring, in that it is composed mainly of small particles (less than 100 micrometres across) that must be constantly replenished from larger source bodies such as moons, because such tiny grains are eroded away or ejected from their host planet's orbit on very

short timescales (less than 1,000 years). The Phoebe ring, like Jupiter's gossamer rings, probably consists of the dust ejected from moon surfaces by impacts²; by contrast, Saturn's E ring is supplied by geysers emanating from the interior of the planet's moon Enceladus.

But the Phoebe ring is vastly different in scale from other dusty rings. It has a core radius (the distance from Saturn at which the ring's density reaches its peak value) that is about 200 times the radius of Saturn and 50 times that of the E ring, the previous record holder for the Solar System's largest planetary ring. Thanks to its huge dimensions, the Phoebe ring is at least ten times more massive than the E ring, despite having a particle number density (20 particles per cubic kilometre) that is tens of millions of times lower¹.

The destiny of all that mass may be the most interesting aspect of the Phoebe ring's discovery. Just as the icy dust in the E ring spreads out from Enceladus and seems to brightly coat the surfaces of Saturn's inner moons³, so the dust

in the Phoebe ring is expected to spiral inward towards Saturn, where much of it would be swept up by Iapetus, the outermost of Saturn's large moons, whose surface patterns are a 300-year-old mystery.

Because Iapetus keeps one face always towards its parent planet, like most moons in the Solar System, it also keeps one face (the 'leading hemisphere') directed towards its direction of motion. Iapetus' leading hemisphere is among the darkest surfaces in the Solar System (its albedo — the fraction of sunlight that is reflected back into space — is about 4%), whereas the opposite ('trailing') hemisphere and the poles are quite bright. Surfaces of intermediate brightness, however, are almost entirely absent⁴. It has long been considered plausible that dust from Phoebe is the most likely cause of Iapetus' curious coloration^{5,6}, but broad agreement has been elusive. The spectral match between the two is not exact, and it is unclear whether the differences can be explained by dust deposition at hyper-velocity and subsequent mixing with the native Iapetan surface.

Furthermore, the lack of dark material at Iapetus' poles is unexpected under a simple model of infalling pollution, although this might be explained by a model that includes infall plus subsequent thermal processing. On the other hand, the leading alternative theory — that the dark material somehow comes from within Iapetus — has a difficult time explaining the close alignment of the dark terrain with the leading hemisphere. Verbiscer and colleagues' discovery¹ of a disk of material surrounding Saturn, corresponding to Phoebe's orbit, is strong evidence for an external source for the dark material on Iapetus. However, much work remains to be done to determine the origin and fate of the observed dust.

As it spirals inward, some fraction of the Phoebe ring's dust makes it past Iapetus and continues on towards the next likely targets, the moons Hyperion and Titan. But the chaotic rotation of Hyperion⁷ would cause it to become evenly coated with the ring's dust, not asymmetrically as for Iapetus, and for Titan the infall would be just one more component of its already complex surface chemistry⁸.

Further observations of this enormous ring would be highly desirable to better determine its structure and spectral properties. In particular, complementary photometry (measurement of an object's brightness) in visible light would constrain the size distribution and albedos of its component particles — are they really as dark as their presumed source (Phoebe) and destination (Iapetus' leading hemisphere)? Furthermore, such observations could clarify whether other moons besides Phoebe are supplying this ring with dust. Phoebe is by far the largest of Saturn's distant moons, but this does not necessarily make it the best source of dust. Although bigger moons present larger targets for dust-generating impacts, their increased surface gravity correspondingly holds on to

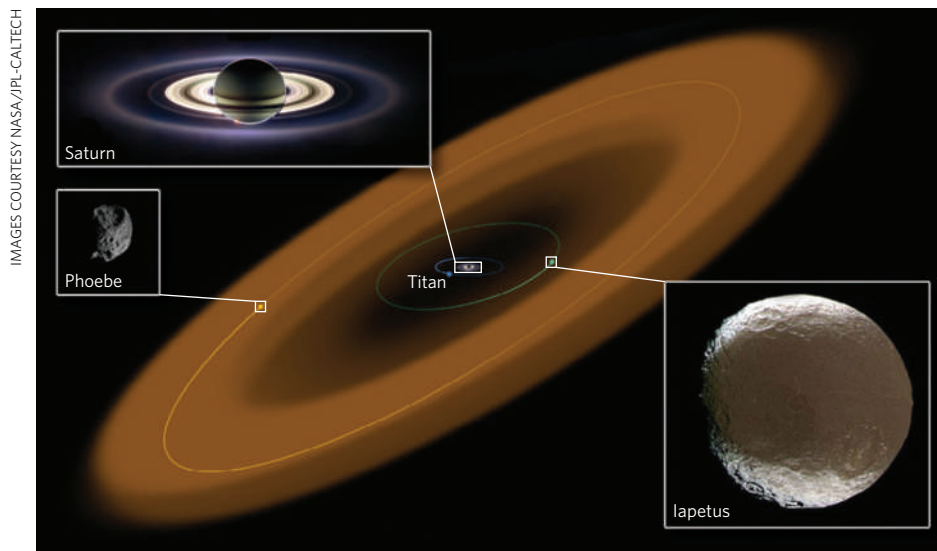


Figure 1 | Saturn's huge ring. This diagram depicts the newly discovered¹ 'Phoebe ring' around Saturn, which spans at least 25 million kilometres and is the largest ring known to be orbiting a planet. The ring corresponds closely to the orbit of Phoebe, the largest of Saturn's outer 'irregular moons', and apparently the source of most of the ring's material. The ring is tilted owing to the influence of the Sun. Dust in the ring probably spirals inward and hits the leading hemisphere of the moon Iapetus, triggering that moon's distinctive two-toned coloration. Also shown are the orbit of Saturn's largest moon Titan, the planet itself and its other rings.

a greater fraction of ejecta, so the best-sized moon for making dust is not obvious².

The data presented by Verbiscer *et al.*¹ provide good evidence that this newly discovered ring does in fact ultimately originate from Phoebe in particular: the ring's vertical profile (seen nearly edge-on from Earth) exhibits a double-layered structure whose thickness corresponds to the vertical distance spanned by Phoebe as it goes along its inclined orbit. This structure is very similar to that of Jupiter's rings and just the sort of profile expected for a population of dust particles originating from Phoebe². It may be, as Verbiscer *et al.*¹ argue,

that Phoebe's contribution has been enhanced by primordial large collisions^{9,10} that generated centimetre-sized and larger chunks of ice and rock that can remain in place over the age of the Solar System. These chunks in turn would generate dust that is ultimately derived from Phoebe and that preserves its orbital characteristics. However, it may well be that Phoebe's little brothers and sisters among Saturn's distant moons also play a significant part in generating the dust that makes up the observed ring, and thus in determining the chemistry of Iapetus.

Whereas nearly everything in the Solar System rotates and orbits in an anticlockwise

direction as seen from above, the Phoebe ring almost certainly rotates backwards ('retrograde'), the first ring known to do so. It is also the first ring known to be significantly tilted (by 27°) to its parent planet's equator. Both of these characteristics are due to the ring's source bodies, Saturn's distant moons, many of which (including Phoebe) have retrograde orbits that are more affected by the Sun than by Saturn's equatorial bulge.

The possibility of giant rings existing around other planets is also worth exploring. Jupiter's moon Himalia is only slightly smaller than Phoebe, and whereas Callisto's coloration is less extreme than Iapetus' in both contrast and distribution, Callisto does have bright and dark terrains, and shows spectral differences between its leading and trailing hemispheres¹¹. Future observations may therefore reveal giant rings at Jupiter and elsewhere, or may show that Saturn's giant ring is as unique as its more famous main rings.

Matthew S. Tiscareno and Matthew M. Hedman are in the Department of Astronomy, Cornell University, Ithaca, New York 14853, USA.
e-mail: matthewt@astro.cornell.edu

1. Verbiscer, A., Skrutskie, M. F. & Hamilton, D. P. *Nature* **461**, 1098–1100 (2009).
2. Burns, J. A. *et al.* *Science* **284**, 1146–1150 (1999).
3. Verbiscer, A. *et al.* *Science* **315**, 815 (2007).
4. Porco, C. C. *et al.* *Science* **307**, 1237–1242 (2005).
5. Soter, S. *IAU Colloq.* **28**, abstr. (1974).
6. Burns, J. A., Hamilton, D. P., Mignard, F. & Soter, S. *Astron. Soc. Pacif. Conf. Ser.* **104B**, 179–182 (1996).
7. Wisdom, J., Peale, S. J. & Mignard, F. *Icarus* **58**, 137–152 (1984).
8. Lunine, J. I. & Lorenz, R. D. *Annu. Rev. Earth Planet. Sci.* **37**, 299–320 (2009).
9. Nesvorný, D., Alvarellos, J. L. A., Dones, L. & Levison, H. F. *Astron. J.* **126**, 398–429 (2003).
10. Turrini, D., Marzari, F. & Tosi, F. *Mon. Not. R. Astron. Soc.* **392**, 455–474 (2009).
11. Moore, J. M. *et al.* in *Jupiter: The Planet, Satellites, and Magnetosphere* (eds Bagenal, F., Dowling, T. E. & McKinnon, W. B.) 397–426 (Cambridge Univ. Press, 2003).

CLIMATE CHANGE

Early survival of Antarctic ice

Damien Lemarchand

Analyses of boron isotopes in ancient marine carbonate sediments provide an enlightening perspective on the links between carbon dioxide and ice-cap cover at a climatically momentous time in Earth's history.

On page 1110 of this issue, Pearson *et al.*¹ report how they have peered back to a time, around 33.5 million years ago, when Earth underwent a drastic interval of climatic change associated with the formation of the Antarctic ice cap. The authors' work provides a fund of data about the connection between climate and atmospheric levels of the greenhouse gas carbon dioxide.

Coupled records of past variations in temperature and atmospheric CO₂ are precious:

they provide fundamental information on climate dynamics that can help in predicting future change. These records are available for recent times from instrumental measurements and, going further back, from data provided by oxygen isotopes in ice cores and by CO₂ trapped in air bubbles in such cores. The ice-core data take us back almost a million years^{2,3}, but that still covers only the last few glacial–interglacial cycles, which were characterized by atmospheric CO₂ concentrations

of 170–300 parts per million by volume (p.p.m.v.) — lower than the present-day value of about 380 p.p.m.v.

To test climate theories in the present-day condition of massive carbon emissions to the atmosphere, what is needed is access to geological archives that provide clues about how past climate responded to CO₂ levels higher than those of today. Analyses of oxygen isotopes in the shells of marine organisms that form carbonate sediments can provide proxy data on ancient global temperature and ice volume, in the same way that oxygen isotopes in ice samples provide such data for more recent times. But, so far, reconstructions of ancient levels of CO₂ have been rare⁴, and they remain much needed.

This is where Pearson *et al.*¹ come in. They have used boron isotopes in exceptionally well-preserved marine microfossils — the carbonate shells of organisms called foraminifera — as an indirect measurement of atmospheric CO₂

ECOLOGY

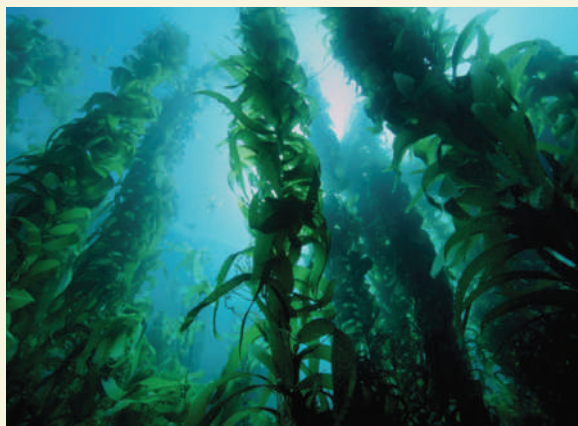
Kelp in postglacial time

In a multifaceted exercise in palaeoecology, Michael Graham and colleagues have taken the subject underwater. Their study is an investigation of how the distribution and productivity of giant kelp have changed since the end of the Last Glacial Maximum, about 20,000 years ago, in one of its present-day strongholds, the Southern California Bight (M. H. Graham *et al.* *Proc. R. Soc. B* doi:10.1098/rspb.2009.1664; 2009).

The significance of giant kelp (*Macrocystis pyrifera*; pictured) is that it is a foundation species — a dominant primary producer — that creates vast submarine forests in coastal areas of temperate regions. These kelp forests constitute ecosystems on which a multitude of other species depend.

Delving into the past of marine algae such as kelp poses particular challenges. Unlike the terrestrial flora, for example, there is no pollen record to track a shift in species abundance. Graham *et al.* had to find less direct approaches.

The authors worked on



the assumption that the environmental needs of kelp have not varied over the past 20 millennia. They created 'niche-based' reconstructions through time of factors that primarily included the availability of a rocky sea floor in adequate light conditions, at depths of up to 25 metres. These reconstructions were complemented by palaeo-oceanographic records that allowed estimates of nutrient availability, and so of kelp productivity.

The result is a millennial-scale account of the dynamics of southern Californian kelp forests. The patterns are

swiftly changing and complex. But the broad picture is of an apparent increase in biomass of three times or so from glacial-maximum levels until about 7,500 years ago, followed by a rapid fall of as much as 70% to today's levels.

An archaeological twist comes from a link between the food resources furnished by kelp forests and early human settlements. There is documentation of this for California, and Graham *et al.* suggest that past human migrations might also have had a kelp connection.

Tim Lincoln

during the Eocene–Oligocene climate transition that began about 34 million years ago. This was a time of dramatic climate perturbation, which saw the formation of the still-extant Antarctic ice cap, and which has been attributed to changes in the global carbon cycle⁵.

Boron has a physicochemical property that is particularly appropriate for reconstructing historical levels of atmospheric CO₂. The relative abundance of its two stable isotopes (¹¹B/¹⁰B, designated $\delta^{11}\text{B}$) in the shells of foraminifera is correlated with the pH of the uppermost ocean layers in which these organisms lived^{6,7}. Because atmospheric CO₂ is in chemical equilibrium with the pH of surface sea water, it is possible to infer past CO₂ concentrations from boron isotopes in carbonate sediments⁸. By comparing their boron-isotope results with published oxygen data⁹ from samples of the same age, Pearson *et al.*¹ reach two intriguing conclusions.

First, they conclude that the slow temperature decline recorded by oxygen isotopes was concomitant with a decline of atmospheric

CO₂ from about 1,100 p.p.m.v. to a threshold concentration of about 750 p.p.m.v., at which the main phase of Antarctic ice-cap growth was initiated. This finding confirms model predictions⁵ that — contrary to what might be expected — the initiation and the rapid expansion of the Antarctic ice sheet occurred about 33.5 million years ago at levels of atmospheric CO₂ that were more than twice the present-day value. Pearson *et al.* propose that the Antarctic glaciation was preconditioned by the global cooling associated with the decline of atmospheric CO₂. But the glaciation really started only when Earth's orbital parameters, which change periodically, favoured the process.

The authors' second conclusion is that, although the newly formed ice cap may have shrunk somewhat, it largely survived a subsequent and rapid recovery of atmospheric CO₂ back to levels of 1,000 p.p.m.v. or more. Such a rise in CO₂ after the main phase of ice-sheet growth is predicted by climate models¹⁰. But the boron isotopes indicate that it occurred within the following 50,000 years, which is

faster than the model prediction. This disparity highlights a need for more refinement in modelling the carbon cycle and understanding its relationship with global climate.

The inferences drawn by Pearson *et al.*¹ about relative variations in boron isotopes — and, hence, in the pH of 'palaeo-seawater' — are solid. But caveats must be mentioned about the extrapolation of boron-isotope data in determining the corresponding atmospheric CO₂ concentration¹¹. Reliable absolute values of seawater pH can be deduced from boron isotopes in shells of ancient marine organisms only once the $\delta^{11}\text{B}$ value of sea water itself is known, which is not readily achieved for the situation 33.5 million years ago. Even when the pH of palaeo-seawater is correctly estimated, more information on seawater chemistry (in particular with respect to the dissolved carbonate species) is still required to deduce the corresponding atmospheric CO₂ at equilibrium. Possible errors in doing so may arise from the method itself — which involves using fluid inclusions in salt deposits to reconstruct open-ocean chemistry¹² — or from the determination of the ancient seawater saturation state with respect to carbonate¹³. The boron data might tell a slightly different story if the model used for palaeo-seawater chemistry turns out to have flaws.

Nonetheless, as they stand, the results validate climate models at CO₂ concentrations not observable in the instrumental and ice-core archives. The unequivocal advance made by Pearson *et al.*¹ is to demonstrate that marine organisms that existed when the Antarctic ice cap formed show much lower $\delta^{11}\text{B}$ values than do such organisms today — probably indicating lower seawater pH and higher CO₂ levels than today — and that they record sharp variations in CO₂ associated with the main phase of ice-cap growth. Their high-quality data will further invigorate study of the coupling between global temperature, ice volume and atmospheric CO₂ in ancient times — and that is no mean achievement. ■

Damien Lemarchand is in the Laboratoire d'Hydrologie et de Géochimie de Strasbourg, UMR 7517 CNRS, EOST/UdS, 67084 Strasbourg Cedex, France.
e-mail: lemarcha@unistra.fr

1. Pearson, P. N., Foster, G. L. & Wade, B. S. *Nature* **461**, 1110–1113 (2009).
2. Lüthi, D. *et al.* *Nature* **453**, 379–382 (2008).
3. Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
4. Pagani, M., Zachos, J. C., Freeman, K. H., Tzippe, B. & Bohaty, S. *Science* **309**, 600–603 (2005).
5. DeConto, R. M. & Pollard, D. *Nature* **421**, 245–249 (2003).
6. Hemming, N. G. & Hanson, G. N. *Geochim. Cosmochim. Acta* **56**, 537–543 (1992).
7. Sanyal, A. *et al.* *Paleoceanography* **11**, 513–517 (1996).
8. Foster, G. L. *Earth Planet. Sci. Lett.* **271**, 254–266 (2008).
9. Zachos, J., Quinn, T. & Salamy, K. *Paleoceanography* **11**, 251–256 (1996).
10. Merico, A. *et al.* *Nature* **452**, 979–982 (2008).
11. Pagani, M., Lemarchand, D., Spivack, A. & Gaillardet, J. *Geochim. Cosmochim. Acta* **69**, 953–961 (2005).
12. Horita, J., Zimmermann, H. & Holland, H. D. *Geochim. Cosmochim. Acta* **66**, 3733–3756 (2002).
13. Rea, D. K. & Lyle, M. W. *Paleoceanography* doi:10.1029/2004PA001064 (2005).

MOLECULAR BIOLOGY

Slip sliding on DNA

Nicholas P. George and James L. Keck

Dedicated binding proteins stabilize single-stranded DNA, protecting it from breakage and distortion. Once thought to form inert complexes with DNA, such proteins are now shown to be remarkably mobile.

The cell's genetic information is maintained as a double helix of interwoven DNA strands in which the information content (the nucleotide sequence) is tucked away between sheltering sugar–phosphate backbones. Although this stable structure benefits long-term information storage, the two strands of DNA must be unwound to allow DNA replication, genetic recombination and repair. Unfortunately, DNA unwinding comes with inherent risks: single-stranded DNA (ssDNA) is susceptible to environmental insults, and is prone to forming aberrant structures that impede subsequent DNA-processing reactions. To alleviate these problems, cellular ssDNA-binding proteins (SSBs) have evolved that bind, protect and stabilize the DNA^{1,2}, providing a platform that allows enzymes to process the single strand. SSBs have long been thought to form relatively immobile complexes with ssDNA. But Roy and colleagues³ report in this issue (page 1092) that an SSB from the bacterium *Escherichia coli* forms a surprisingly dynamic structure with its DNA strand — one in which SSB slides spontaneously and rapidly along ssDNA tracks.

SSBs bind with high affinity and low sequence specificity to ssDNA^{1,2}. Most bacterial SSBs resemble the prototypical *E. coli* protein, which wraps ssDNA around its tetrameric (four-subunit) protein core (Fig. 1). Bacterial SSBs also bind to other proteins involved in DNA metabolism². Eukaryotic SSBs (those of plants and animals), however, have different DNA-binding-domain arrangements and protein-association mechanisms from their bacterial counterparts⁴.

By virtue of the transience of ssDNA species in cellular DNA-processing reactions, SSB–ssDNA complexes have long been thought to be short-lived but essential structures in normal cellular processes. For example, during DNA replication in *E. coli*, the binding protein not only binds to ssDNA, but is also jettisoned from the DNA strand as the DNA template is copied by the cellular replication machinery. This feat requires that multiple SSBs bind to and dissociate from long stretches of ssDNA (in excess of 1,000 nucleotides) within seconds⁵, implying dramatic and rapid dynamics.

Early microscopic and biochemical studies⁶ showed that *E. coli* SSB forms

beaded structures on ssDNA that resemble those of histone proteins (the proteins around which DNA folds to form nucleosomes in the eukaryotic cell nucleus). The movement of histones away from stably bound DNA sequences often requires energy-consuming cellular machinery⁷, and hence one might infer a model for an SSB–ssDNA complex in which there is limited spontaneous sliding of SSB on its DNA. However, kinetic studies of *E. coli* SSB^{8,9} and of a related SSB from a bacterium-infecting virus, bacteriophage T4 (ref. 10), have supported a more dynamic arrangement in which binding proteins are envisaged to move more freely on ssDNA. Thus, the questions of whether SSB moves along its DNA, and how enzymes that process the DNA in SSB–ssDNA complexes might engage their substrates and dislodge the binding protein, have gone unanswered.

In their study, Roy and colleagues³ show that *E. coli* SSB translocates spontaneously and rapidly along the DNA. The finding comes from a set of single-molecule experiments that followed the mobility of the binding protein on ssDNA over time. The protein was estimated to slide spontaneously at a brisk 60 steps per second at 37 °C, moving at a step size of about 3-nucleotide intervals. The authors show that

individual SSB tetramers moved distances in excess of the size of their DNA-binding site (65 nucleotides) in either direction along the ssDNA molecules.

These observations beg the question of how binding-protein dynamics influence enzymes that process the DNA in SSB–ssDNA complexes to maintain the genome. To address this problem, the authors³ examined the relationship between SSB and RecA — an enzyme involved in *E. coli* DNA repair and recombination. RecA forms filaments on ssDNA in a directional manner, and nudges the binding protein along the DNA as the filament elongates. In turn, SSB disrupts hairpin structures in the ssDNA that lie ahead of the growing RecA filament, thereby removing potentially obstructive DNA secondary structures and assisting RecA-filament propagation. Taken together, these studies show how SSB promotes RecA-filament formation through directional diffusion along ssDNA.

The discovery that SSB moves along its DNA partner raises many questions. First, how does SSB diffusion influence the activity of proteins with which it forms complexes? It binds to more than a dozen different proteins and, in many cases, stimulates the activity of its partners².

As *E. coli* RecA does not seem to bind to SSB directly, it is unclear whether the current findings apply to other SSB–ssDNA-processing enzymes. For example, if an enzyme interacts with both SSB and ssDNA, does it bind to only one SSB protein and push downstream SSBs like the carriages of a train, or does it traverse from one binding protein to the next?

Second, how is the binding protein removed from DNA? Considering the high affinity of SSB–ssDNA interactions, this is uncertain. If a long train of ssDNA-bound SSBs is pushed by an SSB–ssDNA-processing enzyme, does the binding protein simply fall off the end of the DNA (as seems to be the case for RecA³), or can it be sequentially removed by the enzyme as SSB–ssDNA complexes are processed (Fig. 1)? In the former model, how is the SSB that is bound to gap structures (ssDNA flanked by double-stranded DNA, as would be found in DNA-replication intermediates) removed without a free DNA end? With the clearer understanding of SSB–ssDNA dynamics provided by Roy and colleagues,

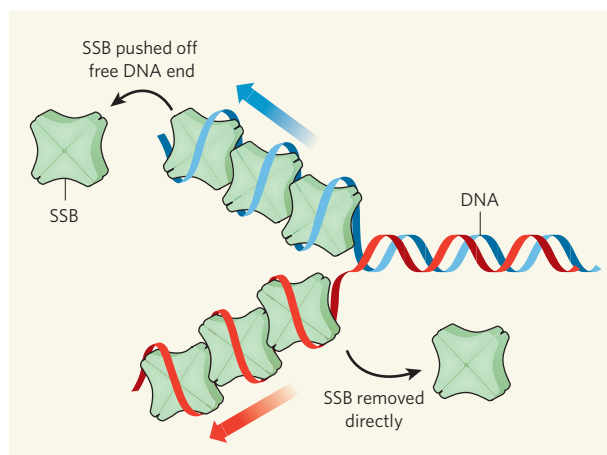


Figure 1 | Binding and removal of SSBs from DNA. Single-stranded DNA-binding proteins (SSBs) bind to single-stranded DNA (ssDNA) by wrapping the single DNA strand around the tetrameric protein core. Roy and colleagues³ report that the binding protein from the bacterium *Escherichia coli* diffuses spontaneously and rapidly along ssDNA. Two possible models for SSB dissociation from DNA are shown. Cellular enzymes that process the DNA in SSB–ssDNA complexes, such as RecA (not shown), might push the binding protein off a free end of ssDNA (top). Alternatively, the protein could be sequentially removed from ssDNA by other enzymes, such as those that bind directly to SSBs, as SSB–ssDNA complexes are processed (bottom). The blue and red arrows indicate the hypothetical direction of movement of the processing enzymes.



50 YEARS AGO

The Scientific Principles of Crop Protection. By Hubert Martin — To say that, nuclear wars apart, the greatest problem of the future will be to feed the rapidly growing human population is none the less true for being trite. Happily there is no need to think it is insoluble. Average crop yields are so low that the scope for improvement is enormous, and starvation can be avoided for a long time simply by improving the health of crops. Over much of the world most crops are left to fend for themselves, unaided in their struggle with pests and diseases. What annual toll these predators take cannot be estimated at all accurately, but there is little doubt that human beings will have at least twice as much to eat when they stop sharing their crops with pests and diseases.

ALSO:

Prof. Jaroslav Heyrovsky ... has been awarded the Nobel Prize for Chemistry for 1959, for his discovery and development of polarography ... The number of papers dealing with polarography now approaches the 10,000 mark, and the technique finds application in many fields of chemistry and biochemistry ... For example, the kinetics of electrode reactions and of chemical reactions associated with redox processes have been studied, redox potentials have been determined and the energetics of the reduction of organic compounds have been elucidated. From *Nature* 24 October 1959.

100 YEARS AGO

Considerable interest attaches to the discovery of large quantities of shells of the pearl-mussel (*Unio margaritifera*) in gravel of apparently Pleistocene age in the Thames near Mortlake ... The cause of the extinction of the species is explained by the fact that as the land sank the river became more sluggish, and silt and mud commenced to accumulate. Such conditions would prove highly detrimental to its welfare, and the species soon ceased to exist. From *Nature* 21 October 1909.

these fundamental questions can begin to be addressed.

Nicholas P. George and James L. Keck are in the Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin 53706, USA. e-mail: jlkeck@wisc.edu

1. Lohman, T. M. & Ferrari, M. E. *Annu. Rev. Biochem.* **63**, 527–570 (1994).
2. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318 (2008).
3. Roy, R., Kozlov, A. G., Lohman, T. M. & Ha, T. *Nature* **461**, 1092–1097 (2009).
4. Wold, M. S. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
5. Kornberg, A. & Baker, T. A. *DNA Replication* 2nd edn (Freeman, 1992).
6. Chrysogelos, S. & Griffith, J. *Proc. Natl Acad. Sci. USA* **79**, 5803–5807 (1982).
7. Clapier, C. R. & Cairns, B. R. *Annu. Rev. Biochem.* **78**, 273–304 (2009).
8. Römer, R. et al. *Biochemistry* **23**, 6132–6137 (1984).
9. Kuznetsov, S. V. et al. *J. Mol. Biol.* **359**, 55–65 (2006).
10. Lohman, T. M. & Kowalczykowski, S. C. *J. Mol. Biol.* **152**, 67–109 (1981).

BIOCHEMISTRY

Enzyme's black box cracked open

David H. Sherman

Polyketide synthase enzymes make compounds from molecules that synthetic chemists can't easily control. The basis of the enzymes' ability to use such unstable precursors has been laid bare.

Polyketides form a structurally diverse — and often medicinally useful — family of compounds derived from bacteria, fungi and plants. These compounds include the antibiotic erythromycin, the anticancer agent epothilone and the cholesterol-lowering drug lovastatin. But it's not all good news. Some polyketides are dangerous, such as aflatoxin B₁, which causes liver cancer in humans. The biological activities of polyketides are prescribed by their molecular shapes and sizes. These, in turn, are ultimately determined by the biosynthetic machinery that makes the molecules — polyketide synthase (PKS) enzymes, along with other auxiliary enzymes that tailor polyketide structures to provide additional chemical diversity¹.

Among PKS enzymes, some of the most confounding and mysterious are the fungal non-reducing, multi-domain iterative PKSS (NR IPKSs). These enzymes perform a truly remarkable 'ring-closing' transformation: they convert linear polyketide fragments into a series of connected aromatic rings, known as polycycles. On page 1139 of this issue, Crawford *et al.*² provide stunning insight into how NR IPKSs bind and fold polyketides to control the structural outcome of these transformations. They have obtained a crystal structure of a domain of PksA — an NR IPKS found in certain species of *Aspergillus* moulds that is involved in the assembly of a polycyclic precursor to aflatoxin B₁, known as noranthrone.

The stage was set for Crawford and colleagues' work when a region of PksA known as the product-template (PT) domain was reported last year to be responsible for catalysing polyketide ring-closing reactions³. This unique functional domain was shown to be evolutionarily conserved in NR IPKSs from diverse fungi. But how does it work? The mechanism might reasonably be expected to be similar to that of fatty acid synthase (FAS)

enzymes, which are phylogenetically and biochemically related to NR IPKSs. FAS enzymes make the long hydrocarbon chains of saturated fatty acids by repeating a cycle of reactions, in which the chain is first elongated by the addition of a small molecular building block and then chemically reduced⁴.

In fact, although PksA makes molecular chains using elongation steps similar to those of FAS enzymes, it lacks the processing machinery necessary to perform reduction reactions. This results in the formation of a polyketide chain that, in isolation, would be so chemically unstable as to be practically useless. But in NR IPKS enzymes, the polyketide is covalently tethered to an arm (known as the phosphopantetheine arm) of one of the protein domains that makes up the enzyme. The combination of the covalent linkage and the NR IPKS protein scaffold provides a protective environment that stabilizes the polyketide.

The importance of chain stabilization in polyketide biosynthesis was suggested several decades ago when a series of linear polyketides was prepared by chemical synthesis⁵. The compounds were made to see how they might be involved in the biosynthesis of natural products that are polycyclic and aromatic, many of which are derived from fungi. These studies showed that, in the absence of external influences, polyketides undergo several different uncontrolled cyclization reactions (rather than just one), yielding a mixture of products (Fig. 1a). This ultimately suggested that NR IPKS enzymes are responsible for much more than just chemical catalysis — they must also direct the polyketide chain toward a pre-cyclization configuration, something that was unimaginable at the time. Such molecular control would enable faithful generation of a single reaction product out of a myriad of possibilities.

The 'black box' of fungal NR IPKS has now

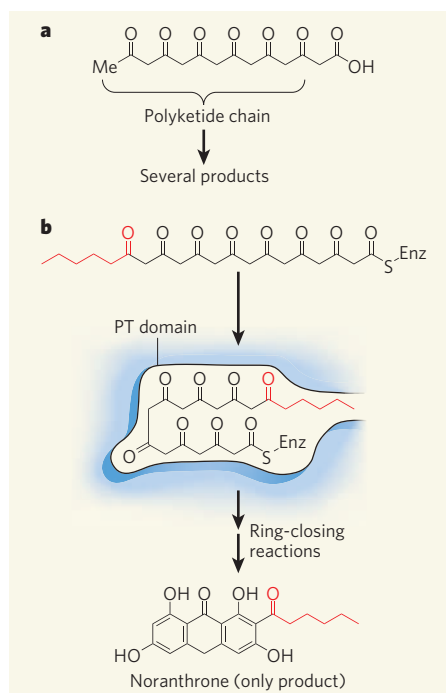


Figure 1 | Controlling the reactivity of polyketide chains. **a**, Molecules containing polyketide chains are highly unstable. In the absence of external control, they undergo several reactions, yielding a mixture of products. **b**, Fungal non-reducing, multi-domain iterative polyketide synthase (NR IPKS) enzymes make polyketide-containing compounds as precursors of natural products. The polyketide chains are covalently linked to the enzyme (Enz) and may undergo further modification — in this case, a hexanoyl group (red) is attached at one end of the chain. The resulting product is then guided into the product-template (PT) domain of the enzyme, which folds the polyketide chain into the optimal conformation for subsequent ring-closing reactions. Crawford *et al.*² have obtained crystal structures of the PT domain of a fungal NR IPKS, revealing the mechanisms used by the enzyme to control its substrate, which allow it to faithfully form a single compound, noranthrone, rather than a mixture of products. Me, methyl group.

been cracked wide open in Crawford and colleagues' penetrating account² of the pivotal role of the PT domain. First, the authors find that, although the amino-acid sequence of PksA bears little resemblance to those of FAS enzymes, the crystal structure of the PT domain of PksA has a similar 'double hotdog fold'⁶ to that of certain domains of FAS enzymes — specifically, the dehydrase domains that form part of the FAS machinery responsible for reductive processing reactions. This indicates that the PT domain has evolved from a FAS dehydrase domain, or from its counterpart in bacterial modular PKSs.

The authors also observe a clear binding region for the phosphopantetheine arm of PksA that guides polyketide substrates into the PT domain's binding pocket (Fig. 1b). Furthermore, by obtaining the crystal structure of PksA in which a model substrate is bound,

Crawford *et al.* reveal the presence of a spacious cyclization chamber that can harbour two aromatic rings. Herein lie the two amino-acid residues that initiate the key ring-closing reactions catalysed by the PT domain. Finally, the authors' crystal structures strongly suggest that the binding pocket orientates the polyketide chain ready for catalysis by anchoring the chain's two termini at opposite ends of the pocket.

Crawford and colleagues' analysis² of PT domains from other NR IPKS systems reveals that, within an otherwise highly conserved amino-acid sequence, the most significant level of divergence occurs within the cyclization chamber. This is consistent with the idea that the chambers of different enzymes control the overall configurations of specific substrates prior to reaction, and that they have each evolved to accommodate the correct number of ring-forming reactions necessary to generate a specific product. Taken together, the authors' findings show that the PT domain is the main determinant of chemical architecture produced from NR-IPKS biosynthetic pathways. After the cyclization reaction, the final release of the polyketide product is catalysed by the terminal domain of NR IPKS enzymes, a process that

often results in the formation of another ring in the natural product.

The data presented by Crawford *et al.*², along with those from a recent study of another previously elusive enzyme from a bacterial PKS⁷, bring new clarity to the once murky chemical world of polyketide ring cyclization in natural-product assembly. This fresh insight will certainly provide exciting opportunities for engineering metabolic systems for the preparation of biologically active natural products that have great potential as medicinal agents. ■ David H. Sherman is at the Life Sciences Institute and Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109-2216, USA.

e-mail: davidhs@umich.edu

1. Wilkinson, B. & Micklefield, J. *Nature Chem. Biol.* **3**, 379–386 (2007).
2. Crawford, J. M. *et al.* *Nature* **461**, 1139–1143 (2009).
3. Crawford, J. M. *et al.* *Science* **320**, 243–246 (2008).
4. Leibundgut, M., Maier, T., Jenni, S. & Ban, N. *Curr. Opin. Struct. Biol.* **18**, 714–725 (2008).
5. Harris, T. M. & Hay, J. V. *J. Am. Chem. Soc.* **99**, 1631–1637 (1977).
6. Dillon, S. C. & Bateman, A. *BMC Bioinformatics* **5**, 109 (2004).
7. Ames, B. D. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 5349–5354 (2008).

MICROSCOPY

Light from the dark

Stefan W. Hell and Eva Rittweger

Fluorescence microscopy is the most popular way to image biomolecules, but it leaves many of them in the dark. Non-fluorescent, light-absorbing molecules can now be viewed by a method that turns them into mini-lasers.

What happens when a molecule absorbs a photon from a beam of light? It moves from the ground state to an excited, higher-energy state and then quickly relaxes, giving off the absorbed energy as heat. There are, however, notable exceptions — molecules called fluorophores, which, after some 'wiggling', relax by emitting a lower-energy photon. Employed as molecular tags, fluorophores are invaluable in biomedical microscopy and diagnostics because they render dark molecules visible with high specificity. But what if fluorescent tagging is not an option, as in applications such as endoscopy, and the molecules under investigation stay hidden in the dark? As Min *et al.*¹ report on page 1105, it is still possible to squeeze photons out of such molecules to produce three-dimensional images of biological systems, such as living cells and tissues.

How is this possible? Well, another mechanism for molecular relaxation exists that can be induced by a beam of light. In this process, called stimulated emission, a photon encountering an excited molecule produces a copy of itself, thus adding another photon of the same colour and propagation direction to the beam.

To be effective, the energy of the stimulating photon must match the gap between the excited and the ground state. In fact, the stimulating photons need to be slightly lower in energy than their excitation counterparts, because some of the excited molecule's energy is usually lost as vibrational motion (wiggling) before the photon arrives². Stimulated emission is used to amplify light in lasers³, and to overcome the resolution barrier in fluorescence microscopy⁴. As a molecular process it is almost as effective as light absorption, because both processes depend on the molecule's photon-capture area of about 0.2×0.2 nanometres, which is roughly the area of the molecule itself.

The role of photon-capture area in light absorption and stimulated emission can be thought of as follows. Imagine two synchronized trains of laser pulses directly focused on a molecule: pulses containing excitation photons are followed by pulses of photons for stimulated emission (Fig. 1), with each pulse containing N photons. If one could produce focal light spots that are the size of the molecule, then every photon in the pulses would interact with the molecule — a single photon

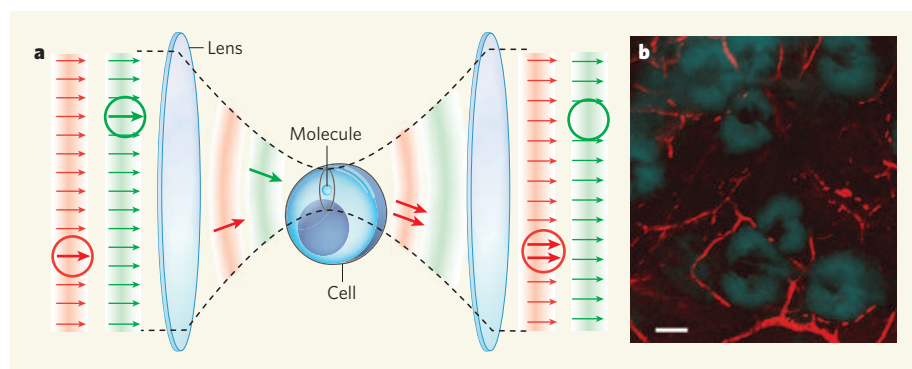


Figure 1 | Imaging non-fluorescent molecules using stimulated emission. **a**, In Min and colleagues' imaging method¹, molecules in a biological specimen are irradiated by ultrashort pulses of light (green). If they absorb a photon, the molecules enter an excited state. The excited molecules are then irradiated by another pulse of light of slightly lower energy (red). When a photon from the second pulse interacts with an excited molecule, the molecule relaxes back to its ground state by emitting a duplicate photon. Overall, a photon from the excitation pulse is absorbed and so is lost (green circles), whereas a photon from the second pulse is doubled up, adding a new, identical photon to the transmitted light (red circles). The gain in photons is measured to quantify the number of molecules in the sample. The colours shown do not correspond to the actual wavelengths of light used in the experiments. **b**, Using their technique, the authors obtained this image of microcapillaries in *ex vivo* mouse skin. The red parts of the image are caused by stimulated emission from haemoglobin in red blood cells. These are overlaid with an image taken using transmitted light, showing sebaceous glands. Scale bar, 50 μm . (Image courtesy of W. Min, S. Lu and S. Chong, Harvard University.)

($N = 1$) from the first pulse would be absorbed and excite the molecule, and another from the second pulse would duplicate itself immediately afterwards.

Unfortunately, because of diffraction, the pulses cannot be focused on a region smaller than about 200×200 nanometres. This area is more than a million times larger than that of a molecule, and so a single photon would most probably miss its target. If, however, more than 1 million photons are used per pulse, one of them will certainly do the job, albeit at the cost of millions of surplus photons that have to be discarded. Discarding surplus excitation photons from fluorescence photons is easily done using filters, and is routine in fluorescence imaging; but singling out the duplicate photons in a stimulating pulse is impossible. Fortunately, to detect the presence of a molecule using stimulated emission, it is sufficient to measure only the number of photons that are added to the pulse — but even that is not straightforward, because N fluctuates from pulse to pulse.

Min *et al.*¹ overcome the challenge of gain measurement by rapidly modulating the excitation beam (and thus the production of duplicate photons), and synchronizing the detection of the slightly strengthened stimulating beam with the modulation^{5,6}. Thus, they reduce the fluctuations in the signal to 'shot noise', which for a train of m recorded pulses amounts to $\sqrt{(N \times m)}$. Because a pulse containing about 1 million stimulating photons yields a new photon from each of n excited molecules in a sample, m such pulses deliver $m \times n$ new photons. The increase in signal should become detectable if $m \times n$ is greater than $\sqrt{(N \times m)}$, which implies that $m > N/n^2$ pulses will make a group of n dark molecules visible. In other

words, detecting 20 molecules requires a minimum of 10,000 pulses. The authors' laser system fires about 10^8 pulses per second, which delivers images of molecules for measurement times of approximately 0.2 milliseconds on a single object point — fast enough for imaging.

Much in this new microscopy contrast mode is determined by the lifetime of the excited state, which is less than 1 picosecond (10^{-12} seconds). The excitation and stimulated-emission photons are preferably squeezed into pulses of about 0.2 ps duration (each providing transient intensities of less than 10 GW cm^{-2} in the sample, which is tolerable for living samples), with the stimulated-emission pulses immediately following their excitation counterparts. This ensures that molecules do not relax before the stimulating photons arrive. Because each round is completed in less than 1 ps, the molecules are almost instantly ready for another round. Indeed, the pulse rate produced by Min and colleagues' laser system is thousands of times slower than the maximum rate that could in principle be used. The recording time required for each measurement could therefore be dramatically reduced by firing pulses at higher rates, which should be facilitated by future developments in laser technology. Increased pulse rates might, however, damage the sample being studied, so the irradiation dose would have to remain at a level compatible with (live) cell imaging.

Another method for imaging non-fluorescent molecules has previously been reported⁵, in which the loss of photons, rather than the gain, is measured when an excited state of the molecules absorbs photons to enter an even higher energy state. The problem with this is that molecules in excited states are reactive, which makes them prone to decomposition.

A strong advantage of Min and colleagues' approach¹ is that the excited molecules are always forced back to the non-reactive ground state. Nevertheless, both methods provide three-dimensional resolution⁷ because their signals stem mainly from molecules in the focal region, which can be raster-scanned through a sample to build up an image. Min and colleagues' method¹ also yields more photons per molecule than stimulated Raman scattering, a phenomenon that has recently also been pioneered by the same group⁶ for imaging non-fluorescent molecules.

But how does the new method compare with fluorescence? Unlike stimulated emission, fluorescence is randomly emitted in space and is thus harder to collect. But for most practical purposes, fluorescence still wins out because the signal can be freed from background noise. However, if non-fluorescent molecules are to be studied and fluorescent labelling is impractical, as is the case in many applications, then the stimulated-emission technique will come into its own.

Stimulated emission has been prominently used in stimulated emission depletion (STED) microscopy^{4,8}, to keep all the fluorophores in a sample dark except for those in a spot smaller than the diffraction resolution limit of the microscope. But in STED microscopy, apart from its ability to switch fluorophores off, there is no interest in stimulated emission per se. Indeed, in related nanoscopy methods⁸, stimulated emission has been replaced by other mechanisms for keeping fluorophores dark, such as flipping the spin of one of their electrons (triplet-state pumping) or relocating some of their atoms (*cis-trans* photoisomerization)⁸. By contrast, in Min and colleagues' technique¹, stimulated emission is the actual goal. An intriguing possibility for the future would be to design a set of laser pulses that fulfil both roles of stimulated emission — switching off molecular signals and stimulating photon emission — to provide images of unlabelled, non-fluorescent molecules at sub-diffraction (nanoscale) resolution for the first time. So, for many reasons, Min and colleagues' method is a bold step towards unveiling details of live cells and tissues that would otherwise be left uncharted. ■

Stefan W. Hell and Eva Rittweger are in the Department of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry, 37070 Göttingen, Germany. Stefan W. Hell is also at the German Cancer Research Center (DKFZ), Heidelberg.
e-mail: hell@nanoscopy.de

1. Min, W. *et al.* *Nature* **461**, 1105–1109 (2009).
2. Schäfer, F. P. *Dye Lasers* (Springer, 1973).
3. Schawlow, A. L. & Townes, C. H. *Phys. Rev.* **112**, 1940–1949 (1958).
4. Hell, S. W. & Wichmann, J. *Opt. Lett.* **19**, 780–782 (1994).
5. Fu, D. *et al.* *J. Biomed. Opt.* **12**, 054004 (2007).
6. Freudiger, C. W. *et al.* *Science* **322**, 1857–1861 (2008).
7. Dong, C. Y., So, P. T. C., Buehler, C. & Gratton, E. *Optik* **106**, 7–14 (1997).
8. Hell, S. W. *Nature Biotechnol.* **21**, 1347–1355 (2003).

The DNA-damage response in human biology and disease

Stephen P. Jackson¹ & Jiri Bartek²

The prime objective for every life form is to deliver its genetic material, intact and unchanged, to the next generation. This must be achieved despite constant assaults by endogenous and environmental agents on the DNA. To counter this threat, life has evolved several systems to detect DNA damage, signal its presence and mediate its repair. Such responses, which have an impact on a wide range of cellular events, are biologically significant because they prevent diverse human diseases. Our improving understanding of DNA-damage responses is providing new avenues for disease management.

Each of the $\sim 10^{13}$ cells in the human body receives tens of thousands of DNA lesions per day¹. These lesions can block genome replication and transcription, and if they are not repaired or are repaired incorrectly, they lead to mutations or wider-scale genome aberrations that threaten cell or organism viability. Some DNA aberrations arise via physiological processes, such as DNA mismatches occasionally introduced during DNA replication and DNA strand breaks caused by abortive topoisomerase I and topoisomerase II activity. In addition, hydrolytic reactions and non-enzymatic methylations generate thousands of DNA-base lesions per cell per day. DNA damage is also produced by reactive-oxygen compounds arising as by-products from oxidative respiration or through redox-cycling events involving environmental toxic agents and Fenton reactions mediated by heavy metals². Reactive oxygen and nitrogen compounds are also produced by macrophages and neutrophils at sites of inflammation and infections³. Such chemicals can attack DNA, leading to adducts that impair base pairing and/or block DNA replication and transcription, base loss, or DNA single-strand breaks (SSBs). Furthermore, when two SSBs arise in close proximity, or when the DNA-replication apparatus encounters a SSB or certain other lesions, double-strand breaks (DSBs) are formed. Although DSBs do not occur as frequently as the other lesions listed above, they are difficult to repair and extremely toxic⁴.

The most pervasive environmental DNA-damaging agent is ultraviolet light. Although the ozone layer absorbs the most dangerous part of the solar ultraviolet spectrum (ultraviolet C), residual ultraviolet A and ultraviolet B in strong sunlight can induce $\sim 100,000$ lesions per exposed cell per hour. Ionizing radiation also generates various forms of DNA damage, the most toxic of these being DSBs⁵. Some ionizing radiation results from radioactive decay of naturally occurring radioactive compounds. For example, uranium decay produces radioactive radon gas that accumulates in some homes and contributes to lung-cancer incidence. Exposure to natural or man-made radioisotopes also occurs during cancer radiotherapy, whereas the radioactive compounds iodine-131 and technetium-99m are exploited to diagnose and treat benign and malignant thyroid diseases. Lessons about the health consequences of excessive radiation exposure are provided by the aftermaths of the Chernobyl nuclear-reactor disaster and nuclear detonations over Japan in the Second World War.

Today, probably the most prevalent environmental cancer-causing chemicals are those produced by tobacco products, which trigger

various cancers, most notably those of the lung, oral cavity and adjacent tissues^{6,7}. Cancer-causing DNA-damaging chemicals can also contaminate foods, such as aflatoxins found in contaminated peanuts and heterocyclic amines in over-cooked meats⁷. DNA-damaging chemicals have also been used in warfare, and on a more positive note, are widely used to treat cancer⁸ and ailments such as psoriasis⁹.

Here, we describe how DNA lesions are dealt with at the molecular level. We then explain how such responses affect many cellular processes, their biological significance and their roles in preventing human diseases. Finally, we illustrate how our increasing knowledge of DNA-damage responses is providing opportunities for improving disease detection and management.

An integrated signalling and genome-maintenance network

To combat threats posed by DNA damage, cells have evolved mechanisms—collectively termed the DNA-damage response (DDR)—to detect DNA lesions, signal their presence and promote their repair^{10–12}. Cells defective in these mechanisms generally display heightened sensitivity towards DNA-damaging agents and, as described below, many such defects cause human disease. Although responses differ for different classes of DNA lesions, they usually occur by a common general programme (Fig. 1). Although we focus on DNA repair and DNA-damage signalling separately, we stress that these operate collectively and share many components.

DNA-repair pathways. The wide diversity of DNA-lesion types necessitates multiple, largely distinct DNA-repair mechanisms (Table 1). Whereas some lesions are subject to direct protein-mediated reversal, most are repaired by a sequence of catalytic events mediated by multiple proteins. In mismatch repair, detection of mismatches and insertion/deletion loops triggers a single-strand incision that is then acted upon by nuclease, polymerase and ligase enzymes¹³. In base-excision repair, a damaged base is often recognized by a DNA glycosylase enzyme that mediates base removal before nuclease, polymerase and ligase proteins complete the repair¹⁴ in processes overlapping with those used in SSB repair^{15,16}. The nucleotide excision repair (NER) system, which recognizes helix-distorting base lesions, operates via two sub-pathways that differ in the mechanism of lesion recognition: transcription-coupled NER, which specifically targets lesions that block transcription, and global-genome NER¹⁵. A key aspect of NER is that the damage is excised as a 22–30-base oligonucleotide, producing single-stranded DNA (ssDNA) that is acted upon by DNA polymerases and associated factors before ligation ensues¹⁵.

¹The Gurdon Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ²Danish Cancer Society, Centre for Genotoxic Stress Research, Strandboulevarden 49, DK-2100 Copenhagen, Denmark, and Institute of Molecular Genetics, CZ-14220 Prague, Czech Republic.

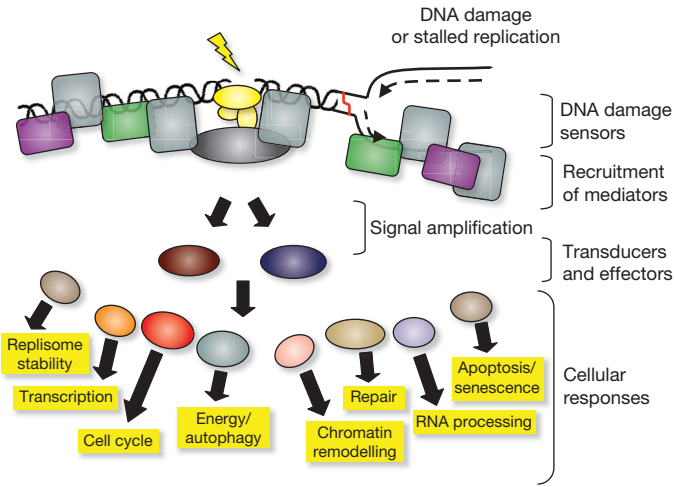


Figure 1 | Model for the DDR. The presence of a lesion in the DNA, which can lead to replication stalling, is recognized by various sensor proteins. These sensors initiate signalling pathways that have an impact on a wide variety of cellular processes. See text for details.

Notably, some DNA lesions are not repaired but are instead bypassed during DNA replication by polymerases with less stringent base-pairing requirements than replicative polymerases¹⁷. For DSB repair, two principal mechanisms are used: non-homologous end-joining (NHEJ)¹⁸ and homologous recombination (HR)¹⁹. In NHEJ, DSBs are recognized by the Ku protein that then binds and activates the protein kinase DNA-PKcs, leading to recruitment and activation of end-processing enzymes, polymerases and DNA ligase IV. A less-well-characterized Ku-independent NHEJ pathway, called microhomology-mediated end-joining (MMEJ) or alternative end-joining, also exists; it always results in sequence deletions²⁰. Although both NHEJ and MMEJ are error-prone, they can operate in any phase of the cell cycle. By contrast, HR is generally restricted to S and G2 because it uses sister-chromatid sequences as the template to mediate faithful repair. Although there are several HR sub-pathways¹⁹, HR is always initiated by ssDNA generation, which is promoted by various proteins including the MRE11–RAD50–NBS1

(MRN) complex. In events catalysed by RAD51 and the breast-cancer susceptibility proteins BRCA1 and BRCA2, the ssDNA then invades the undamaged template and, following the actions of polymerases, nucleases, helicases and other components, DNA ligation and substrate resolution occur. HR is also used to restart stalled replication forks and to repair interstrand DNA crosslinks, the repair of which also involves the Fanconi anaemia protein complex²¹. **DNA-damage signalling and cell-cycle checkpoints.** Key DDR-signalling components in mammalian cells are the protein kinases ATM and ATR, which are recruited to and activated by DSBs and replication protein A (RPA)-coated ssDNA, respectively (Table 1)^{22–24}. Two of the best studied ATM/ATR targets are the protein kinases CHK1 and CHK2 which, together with ATM and ATR, act to reduce cyclin-dependent kinase (CDK) activity by various mechanisms, some of which are mediated by activation of the p53 transcription factor^{23,25,26}. Inhibition of CDKs slows down or arrests cell-cycle progression at the G1–S, intra-S and G2–M ‘cell-cycle checkpoints’, which is thought to increase the time available for DNA repair before replication or mitosis ensues. In parallel, ATM/ATR signalling enhances repair by inducing DNA-repair proteins transcriptionally or post-transcriptionally; by recruiting repair factors to the damage; and by activating DNA-repair proteins by modulating their phosphorylation, acetylation, ubiquitylation or SUMOylation²⁷. Proteomics studies have recently identified a great many as-yet uncharacterized ATM/ATR-mediated phosphorylation sites, suggesting that the DDR modulates additional cellular processes²⁸. If the above events allow effective DNA repair, DDR inactivation ensues, allowing the resumption of normal cell functioning. Alternatively, if the damage cannot be removed, chronic DDR signalling triggers cell death by apoptosis or cellular senescence (that is, permanent cell-cycle withdrawal), both of which have potential antitumour functions^{29,30}. As discussed in later sections, another important SSB- and DSB-signalling protein is the enzyme poly(ADP-ribose) polymerase (PARP). It is becoming increasingly clear that chromatin structure has an impact on the DDR and is modulated in response to DNA damage^{23,31}. The best characterized example of this is ATM/ATR/DNA-PK-mediated phosphorylation of serine 139 of the histone H2A variant, H2AX, on chromatin flanking DSB sites. This brings about ubiquitin-adduct formation in such regions and the recruitment of DDR factors

Table 1 | DDR mechanisms and components

DDR mechanism	Prime lesions acted upon	Key protein components
Direct lesion reversal	O ⁶ alkylguanine	O ⁶ -methylguanine methyltransferase
Mismatch repair	DNA mismatches and insertion/deletion loops arising from DNA replication	Sensors MSH2–MSH6 and MSH2–MSH3 plus MLH1–PMS2, MLH1–PMS1, PLH1–MLH3, EXO1, polymerases δ and ε, PCNA, RFC, RPA, ligase I
Base excision repair and SSB repair	Abnormal DNA bases, simple base adducts, SSBs generated as base-excision repair intermediates by oxidative damage or by abortive topoisomerase I activity	DNA glycosylases (sensors), APE1 endonuclease, DNA polymerases (β, δ, ε) and associated factors, flap endonuclease FEN1, ligase I or ligase III. SSB repair can also involve polymerase β lyase activity, XRCC1, PARP1, PARP2, polynucleotide kinase and aprataxin.
Nucleotide excision repair (NER)	Lesions that disrupt the DNA double helix, such as bulky base adducts and ultraviolet photo-products	Sensors elongating RNA polymerase, XPC–HR23B and DDB1/2, plus XPA, XPE, XPF/ERCC1, XPG, CSA, CSB, TFIIH (containing helicases XPB and XPD), DNA polymerases including polymerase κ and associated factors, PCNA, RPA, ligase I and III
Trans-lesion bypass mechanisms	Base damage blocking replication-fork progression	‘Error-prone’ DNA polymerases, including polymerases η, ι, κ, REV3 and REV1; plus associated factors
Non-homologous end-joining (NHEJ)	Radiation- or chemically-induced DSBs plus V(D)J and class-switch recombination intermediates	Sensors Ku and DNA-PKcs plus XRCC4, XLF/Cernunnos and ligase IV. Can also use the MRE11–RAD50–NBS1 complex, Artemis nuclease, polynucleotide kinase, aprataxin and polymerases μ and λ.
Homologous recombination (HR)	DSBs, stalled replication forks, inter-strand DNA crosslinks and sites of meiotic recombination and abortive topoisomerase II action	RAD51, RAD51-related proteins (XRCC2, XRCC3, RAD51B, RAD51C, RAD51D, DMC1), RAD52, RAD54, BRCA2, RPA, FEN1, DNA polymerase and associated factors. Promoted by MRN, CtIP, BRCA1 and the ATM signalling pathway.
Fanconi anaemia (FANC) pathway	Inter-strand DNA crosslinks	FANCA, FANCC, FANCD1/BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FANCN plus factors including PALB2 and HR factors
ATM-mediated DDR signalling	DSBs	ATM, MRN and CHK2. Promoted by mediator proteins such as MDC1, 53BP1 MCPH1/BRIT1, and by ubiquitin ligases RNF8, RNF168/RIDDLIN and BRCA1.
ATR-mediated DDR signalling	ssDNA, resected DSBs	Sensors ATR, ATRIP and RPA plus the RAD9–RAD1–HUS1 (9-1-1) complex, RAD17 (RFC1-like) and CHK1. Promoted by MRN, CtIP and mediator proteins such as TOPBP1, Claspin, MCPH1/BRIT1 and BRCA1.

See text for details.

plus other chromatin-modifying components which, together, are thought to promote DSB repair and amplify DSB signalling²⁷. Notably, ATM activation leads to chromatin relaxation at sites of DSBs³², and H2AX tyrosine 142 phosphorylation was recently shown to function in the DDR^{33,34}. It therefore seems likely that further DDR-induced chromatin modifications await discovery.

DDR events operate in diverse biological settings

Generating immune-receptor diversity. The only known programmed genome alterations in vertebrates are V(D)J recombination, class-switch recombination and somatic hyper-mutation^{35,36}. These occur in developing B and T lymphocytes to generate immunoglobulin and T-cell receptor (TCR) diversity, thus allowing effective recognition of diverse pathogens and antigens. Immunoglobulin and TCR proteins comprise variable regions that specify antigen binding, and constant regions that endow specific properties to the TCR or the various immunoglobulin classes. Exons encoding the antigen-binding portions of these molecules are composed of V, D and J segments that are combined in various ways to generate mature immunoglobulin and TCR genes. Each segment is flanked by recombination-signal sequences that are recognized by the RAG1–RAG2 protein complex, which generates a blunt DSB at the signal sequence and a covalently-closed DNA hairpin at the coding end. These structures are then processed and ligated by the NHEJ apparatus³⁵. Consequently, besides causing ionizing radiation hypersensitivity, NHEJ defects yield severe-combined immune-deficiency.

A rearranged immunoglobulin heavy-chain variable domain is initially expressed fused to an Igμ constant region but, during antigen-stimulated B-cell differentiation, class-switch recombination can juxtapose a V region to any of several constant regions that bestow distinct properties on the encoded immunoglobulin. B cells undergoing antigen stimulation also activate somatic hyper-mutation to increase mutation rates in the heavy- and light-chain V regions, thus expanding the repertoire of variable segments and allowing selection of B cells expressing immunoglobulin molecules with heightened antigen affinity. Unlike V(D)J recombination, class-switch recombination and somatic hyper-mutation require activation-induced deaminase (AID). AID targeting to variable-region exons and IgH switch regions is believed to trigger deamination of cytosine to uracil, resulting in U•G mismatches that are processed by mismatch repair and/or base-excision repair to yield SSBs. In somatic hyper-mutation, error-prone repair of these SSBs is thought to yield mutations within the variable exon; whereas in class-switch recombination, SSBs are converted into DSBs that are acted upon by NHEJ to juxtapose the immunoglobulin variable exon to a constant-region exon³⁵.

Production of gametes for sexual reproduction. The DDR also has a key role in generating genetic diversity via sexual reproduction, a stage in which is meiosis, the cell-division pathway that generates haploid gametes. After DNA replication, meiosis proceeds by two successive cell divisions: MI that reductionally segregates the two copies of individual chromosomes; and MII that separates resulting sister-chromatid pairs. Before MI, homologous chromosomes align and exchange genetic information by HR³⁷. In species ranging from yeast to human, meiotic

HR is triggered by the topoisomerase-II-related enzyme, Spo11, which generates Spo11-bound DSBs. Spo11 removal and DSB resection then ensue by mechanisms requiring the MRN complex, resulting in ssDNA that promotes HR with the homologous chromosome. These events require all mitotic HR components together with the meiosis-specific RAD51-like protein DMC1³⁷. Consequently, mice deficient in Spo11 or Dmc1 are healthy but infertile. In addition, DDR factors such as ATM, MRN and H2AX monitor and coordinate meiotic HR progression.

Telomere homeostasis and ageing. In most organisms, the ends of chromosomes are organized into telomeres that comprise stretches of short-tandem-DNA repeats terminating in a 3' protruding ssDNA overhang. These repeats are normally generated by the ribonucleoprotein complex telomerase, but in some cancer cells they are maintained by HR-based 'alternative lengthening of telomeres' mechanisms³⁸. Although telomeres possess DNA ends, their sequestration into a complex termed Shelterin prevents them from engaging in NHEJ-mediated fusions or activating ATM/ATR signalling³⁹. Nevertheless, various DDR proteins have important roles at normal telomeres (Table 2); consequently, their defects cause telomere shortening and/or telomere dysfunction that trigger chromosome fusions and ensuing chromosomal instability^{38,40}. Furthermore, mammalian telomeres are recognized by MRN and ATM during G2, possibly to trigger a localized DDR that promotes telomere end-processing and Shelterin-complex formation. HR proteins such as RAD51D are also required for telomere integrity, which might reflect their involvement in establishing T-loop structures wherein the telomeric 3' overhang invades the DNA duplex of internal telomeric sequences^{38–41}.

Excepting specialized cells such as stem cells, human cells generally do not express sufficient telomerase to counteract telomere shortening caused by the inability of the DNA replication machinery to fully replicate chromosomal ends. Thus, human telomeres generally shorten with each cell division until some retain hardly any terminal telomeric repeats. These then fail to act as telomeres and are instead recognized as DSBs, triggering chromosomal fusions and ensuing breakage–fusion–breakage cycles. Under such situations of chronic DDR activation, cells enter into apoptosis or senescence^{40,41}. There is evidence that such processes occur during natural ageing and under circumstances of high cell turnover, such as chronic inflammation or persistent infections. Consistent with ageing in part reflecting the accumulation of shortened telomeres and/or DNA damage, markers of unrepaired DSBs accumulate with age in human and mouse cells, and certain DDR-defective mice strains display hallmarks of accelerated ageing^{42–44}. Furthermore, senescent cells occur at sites of age-related pathologies in man, including atherosclerotic lesions, skin ulcers and arthritic joints^{29,44}.

Physiological control of the DDR. The differing physiologies of various cells presumably impose different DDR requirements. Indeed, some DNA-repair pathways are downregulated upon cell differentiation, possibly reflecting rapid DNA repair being less imperative for non-dividing cells. For example, work on terminally differentiated neurons and macrophages has indicated the existence of a new type of NER, termed differentiation-associated repair, in which both transcribed and non-transcribed DNA strands are

Table 2 | DDR proteins acting at normal telomeres

DDR protein	Role at normal telomeres
MRN	Telomere length regulator, role in end processing
ATM, CHK2 and ATR	Maintain telomere length, phosphorylate Shelterin-complex components, possible roles in telomerase activation and recruitment
Ku and DNA-PKcs	Telomere components and telomere-length regulators, possible telomere-capping functions
RAD9–RAD1–HUS1 (9-1-1)	Telomere component and telomere-length regulator; aids telomerase recruitment/activation
Nucleases EXO1, FEN1, XPF/ERCC1 and Apollo	Processing of telomeric termini to promote telomerase action; regulate telomere integrity
PARP1	Potential telomere-length regulator
BRCA1	Telomere maintenance
RPA	Telomere component, role in telomerase recruitment
WRN	Maintains telomere structure and functions in telomere replication
RAD51D and other HR proteins	Regulate telomeric integrity

See text for details and Table 1 for protein functions.

repaired effectively but non-transcribed loci are repaired poorly or not at all⁴⁵. DSB-repair requirements also change during mammalian nervous-system development, with HR being crucial during neuron proliferation, whereas NHEJ becomes critical as neurons terminally differentiate⁴⁶. This could reflect NHEJ being the prime DSB mechanism available to post-mitotic neurons, whereas dividing neural precursors also use HR. Because of their importance for tissue homeostasis and renewal, it has been speculated that stem cells will rely heavily on the DDR. Indeed, defects in base-excision repair, NER, mismatch repair, HR or the Fanconi anaemia complex impair stem-cell function⁴⁷, and NER or NHEJ defects trigger age-related haematopoietic stem-cell failure in mice^{48,49}.

Many organisms regulate physiological processes in synchrony with the light–dark cycle via the circadian rhythm/biological clock that is controlled by light stimuli. Recent work has established molecular linkages between the biological clock and DDR events⁵⁰. For instance, the *Caenorhabditis elegans* biological-clock gene *clk-2* affects radiation sensitivity, and CLK-2 and its mammalian counterpart control the S-phase checkpoint in response to replication stress⁵¹. Also, it was recently reported that NER is regulated by the circadian clock⁵². Perhaps such linkages allow cells to enhance DDR proficiency at times when physiological or environmentally-induced DNA lesions are most prevalent.

Life cycles of pathogens. The cells of pathogens also possess DDR proteins to mitigate the effects of DNA damage. Furthermore, mutational repair and recombination occasionally occur in viruses, thus fuelling evolution of pathogens such as avian influenza (H5N1) and swine-origin influenza (H1N1) viruses⁵³. In addition, certain pathogens use DDR mechanisms to promote virulence. For example, African trypanosomes—unicellular eukaryotic parasites that infect mammals, including humans—evade immune surveillance by using HR to periodically alter their protective variant-surface-glycoprotein coat⁵⁴. Furthermore, acquisition of drug or pesticide resistance in certain bacteria, plants and unicellular eukaryotic pathogens often involves integration of resistance genes into the organism's genome via DSB-repair mechanisms. DDR activation is also triggered when cells are infected by viruses, including retroviruses (such as HIV-1), adenoviruses, herpes simplex viruses 1 and 2, hepatitis B virus, Epstein–Barr virus, cytomegalovirus, Kaposi's sarcoma virus, simian virus 40 and polyomavirus⁵⁵. Indeed, DDR factors often provide a line of defence against these pathogens, and in many cases viruses have evolved ways to evade such responses. For example, the E6 protein of human papilloma virus types 16 and 18 targets the p53 tumour suppressor for proteolytic degradation to prevent apoptosis of infected cells⁵⁶. Furthermore, the MRN complex and NHEJ components curtail adenovirus infectivity by concatemerizing the viral genome; the virus circumvents this by impairing DNA-PK activity, disrupting complexes containing MRN, and targeting MRN for degradation. Conversely, host cell DDR activities sometimes facilitate viral infectivity. For instance, NHEJ-mediated conversion of linear viral double-stranded (ds) DNA into circles seems to be important for herpes simplex virus replication⁵⁷. Furthermore, CHK2 deficiency or ATM inhibition impairs HSV-1 growth⁵⁸. Retroviruses have dsRNA genomes that are converted into dsDNA, which must then integrate into the host genome to produce new retroviruses. Notably, ATM, MRN and NHEJ proteins are required for efficient retrovirus infection, probably by promoting repair of viral integration intermediates^{55,59}.

The DDR and human disease

Cancer and DNA damage: an intimate relationship. A fundamental feature of cancer is genome instability⁶⁰. For example, genomic instability in lymphoid tumours frequently corresponds to chromosomal translocations, wherein proto-oncogene loci are fused to those of antigen receptors, apparently by aberrant antigen-receptor recombination^{35,36}. In addition, mismatch repair defects cause microsatellite instability that predisposes to colorectal and endometrial

carcinomas¹³. Furthermore, chromosomal instability is seen in most sporadic solid tumours⁶¹. It is likely that transient chromosomal instability arises when telomeres in a nascent tumour become critically short and prone to chromosomal fusions⁶², whereas activated oncogenes and ensuing DNA-replication stress with DSB formation fuel chromosomal instability continuously³⁰. At later stages of cancer progression, chronic hypoxia and/or cycles of hypoxia and re-oxygenation might also contribute to genomic instability and deregulate DDR pathways⁶³.

Most carcinogens operate by generating DNA damage and causing mutations^{15,26}. Furthermore, inherited DDR defects commonly predispose to cancer, contribute to the 'mutator phenotype' of many malignancies, and may allow tumour-cell survival and proliferation despite enhanced mutation rates and genome instability (Supplementary Table 1). Notably, aberrant cell proliferation, caused by oncogene activation or inactivation of certain tumour suppressors, elicits DNA-replication stress and ongoing DNA-damage formation. Such damage activates ATR/ATM-mediated signalling, causing cell death or senescence in cell-culture models and during tumorigenesis *in vivo*^{29,30,64,65}. Indeed, the DDR is commonly activated in early neoplastic lesions and probably protects against malignancy^{64,65}. It has been suggested that breaches to this barrier, arising through mutational or epigenetic inactivation of DDR components, are subsequently selected for during tumour development, thus allowing malignant progression. This model for the DDR as an anticancer barrier helps to explain the high frequency of DDR defects in human cancers³⁰.

Neurodegenerative disorders. Accumulation of DNA lesions in neurons is associated with neurodegenerative disorders, including ataxias together with Alzheimer's, Huntington's and Parkinson's diseases (Supplementary Table 1)^{66,67}. One reason for this may be that neurons generally exhibit high mitochondrial respiration and associated reactive-oxygen-species production that can damage mitochondrial and nuclear DNA⁶⁸. Consistent with a role for base-excision repair and SSB repair in repairing such lesions, defects in these pathways trigger neuronal dysfunction and degeneration^{66,69}. Another reason why the nervous system is particularly vulnerable to DNA damage is its limited capacity for cell replacement in adulthood, potentially leading to accumulation of damaged but irreplaceable terminally differentiated neurons. Furthermore, being in G₀, such cells do not repair DSBs by HR but must use error-prone NHEJ⁶⁶. It is also noteworthy that neurons rely heavily on transcription and that oxidative DNA damage can block this. Thus, accumulation of DNA lesions in repair-defective patients—and possibly in ageing normal individuals—might progressively deprive neurons of vital transcripts, leading to cell dysfunction or apoptosis⁷⁰. Such processes presumably contribute to the neurodegeneration observed in ataxias and in Cockayne syndrome, which are caused by defects in DNA strand-break repair and transcription-coupled NER, respectively^{66,67}.

Genome instability in other heritable human diseases. DNA-repeat instability causes some 40 known diseases that result from expansions or contractions of genetically unstable DNA repeat sequences, usually a trinucleotide motif, within a specific locus for each disease. This instability is thought to arise through the repetitive nature of these regions allowing aberrant DNA-secondary-structure formation during DNA replication or DNA-repair processes^{71,72}. These neuromuscular and neurodegenerative diseases include fragile X syndromes, Friedrich's ataxia, spinocerebellar ataxias, diabetes mellitus type 2, Creutzfeldt–Jakob disease, myotonic dystrophy and Huntington's disease (see Supplementary Table 1 for examples). Analogously, mutations or rearrangements of mitochondrial DNA can lead to impaired mitochondrial function as found in amyotrophic lateral sclerosis, mitochondrial encephalomyopathy, Leigh syndrome, myoclonic epilepsy, Leber's hereditary optic neuropathy, and additional neuropathies and myopathies⁷³.

Immune deficiencies and infertility. Genome rearrangements involving DDR factors occur during immune-system development,

meaning that DDR defects can cause immune deficiency. For instance, mutations in NHEJ factors yield B- and T-cell immune deficiency, whereas ataxia telangiectasia (AT) and Nijmegen breakage syndrome (NBS) patients (defective in ATM and NBS1, respectively) are prone to sometimes-fatal infections, partly due to impaired immunity (class-switch recombination is particularly affected in AT patients). Furthermore, many cancers arising in such conditions are lymphomas and leukaemias of B- and T-cell origin that can result from impaired V(D)J recombination. Human infertility is a significant issue, with ~20% of males in western countries being affected⁷⁴. As meiotic recombination involves DSB generation, it seems likely that certain DDR defects would cause human infertility. Indeed, DDR signalling is readily detectable during human spermatogenesis⁷⁵, and various inherited DDR deficiencies are characterized by infertility or sub-fertility⁷⁴. A significant proportion of human infertility might therefore be caused by DDR deficiencies.

Ageing, stem-cell dysfunction, cardiovascular disease and metabolic syndrome. There is evidence that ageing is in part caused by accumulated DNA damage⁷⁶. First, various endogenously arising DNA lesions accumulate with age in the nuclear and mitochondrial genomes of healthy mammals, including humans^{42,76,77}. This may reflect not only ongoing DNA-damage induction but also declining DNA-repair capacity over time^{68,73}. Second, patients with inherited DDR defects often display features of premature ageing (Supplementary Table 1). Third, work in various organisms has implicated growth hormone and insulin-like growth factor 1 signalling in regulating longevity, and notably, such signalling is downregulated in response to DNA damage⁷⁶. The evolutionary honing of longevity-regulating pathways may serve as an example of antagonistic pleiotropy, where processes selected as advantageous early in life (such as active DDR signalling) because they promote reproduction and prevent tumorigenesis are detrimental later in life because they lead to stem-cell depletion, hence contributing to ageing.

Cell senescence and apoptosis are suspected causes of ageing under conditions where attempted tissue regeneration causes stem-cell exhaustion^{44,78}. Indicative of DNA damage contributing to such episodes, impaired stem-cell function is exhibited in mice with defects in the Fanconi anaemia, NER, mismatch repair or NHEJ pathways^{48,49,79,80}. Furthermore, whereas p53-induced cell death protects against tumorigenesis, pro-apoptotic p53 activity is harmful in settings such as stroke or heart attack^{81,82}. Induction of p53 by oxidative stress and other sources of DNA damage can also affect the development of atherosclerosis, thus providing a link between the DDR and

cardiovascular disease⁸³. Indeed, growing evidence points to human atherosclerosis being characterized by enhanced DNA damage and DDR signalling, leading to senescence of vascular smooth muscle cells and death of other cells to yield atherosclerotic lesions. Modulating ROS production and the DDR therefore represent potential therapeutic opportunities for atherosclerosis.

Metabolic syndrome is a relatively common condition characterized by aberrant glucose metabolism, insulin resistance and atherosclerosis. Interestingly, ATM-defective patients commonly exhibit insulin resistance and glucose intolerance, whereas mice heterozygous or homozygous for ATM mutations display features associated with metabolic syndrome and atherosclerosis^{84,85}. Furthermore, DDR-regulated kinases target multiple substrates involved in glucose metabolism and the insulin–AKT kinase signalling network^{128,84,85}. Thus, although some linkages between the DDR and metabolic syndrome might be indirect, it is possible that the DDR directly modulates certain aspects of energy metabolism and vascular physiology of relevance to metabolic syndrome.

Harnessing DDR knowledge for treating disease

Cancer. Other than surgery, the most prevalent cancer treatments are radiotherapy and chemotherapies that function by generating DNA damage (Table 3). Although such therapies generate dose-limiting toxicities in normal tissues, they are often efficacious. In part, this reflects most cancer cells being DDR-impaired and them proliferating more rapidly than most normal cells (S phase is a particularly vulnerable time for DNA-damage exposure). Nevertheless, DNA repair provides a common mechanism for cancer-therapy resistance. For instance, it has been reported that glioma stem cells display a heightened DDR and are refractory to radiation treatment⁸⁶, thus potentially helping to explain why glioblastoma is difficult to cure (radiation- and chemotherapy-resistance of cancer stem cells might more generally reflect unique properties of their DDR machinery). It has therefore been speculated that DDR inhibition might enhance the effectiveness of radiotherapy and DNA-damaging chemotherapies; and indeed, various DDR-inhibitory drugs are in pre-clinical and clinical development to test this premise^{87,88}. Another possible application for DDR inhibitors is to block apoptotic events, such as those mediated by CHK2 and p53, thus alleviating toxicities to normal tissues.

Many, and possibly all, cancer cells lack one or other aspect of the DDR due to selective pressures operating during tumour evolution (see above). Indeed, reduced or absence of DDR factors correlates, usually positively, with therapeutic outcome (exceptions are defects in p53 and

Table 3 | Examples of DNA-damaging drugs used to treat cancer

Cancer treatment	Types of DNA lesions induced
Radiotherapy and radiomimetics	
Ionizing radiation	SSBs, DSB, base damage
Bleomycin	
Monofunctional alkylators	
Alkylsulphonates	Base damage, replication lesions, bulky DNA adducts
Nitrosourea compounds	
Temozolomide	
Bifunctional alkylators	
Nitrogen mustard	DSBs, DNA crosslinks, replication lesions, bulky DNA adducts
Mitomycin C	
Cisplatin	
Antimetabolites	
5-Fluorouracil	Cytotoxic metabolite, inhibits base pairing leading to base damage and replication lesions
Thiopurines	
Folate analogues	
Topoisomerase inhibitors	
Camptothecins (Topo I)	DSBs, SSBs, replication lesions; anthracyclines also generate free radicals
Etoposide (Topo II)	
Anthracyclines (doxorubicin, epirubicin, daunorubicin) (Topo II)	
Replication inhibitors	
Aphidicolin	DSBs, replication lesions
Hydroxyurea	

See text for details (modified from ref. 87).

other pro-apoptotic proteins, which commonly yield therapy resistance^{82,89}). Because different DNA-repair pathways can overlap in function, and as one pathway can sometimes 'back-up' for defects in another, inhibition of pathways present in a cancer cell should in some cases have a greater impact on the cancer than on normal tissues (Fig. 2a). A paradigm for this is provided by drugs targeting the enzyme PARP1, which binds SSBs and base-excision repair intermediates to facilitate these repair processes. Notably, PARP inhibitors are relatively non-toxic to normal cells but are strikingly cytotoxic towards HR-defective cells, particularly those impaired in BRCA1 or BRCA2 (Fig. 2b)^{90,91}. On the basis of promising phase 1 data, phase 2 trials are currently underway to test PARP1 inhibitors in BRCA-defective breast cancer and ovarian cancer patients (<http://www.cancer.gov/search/ResultsClinicalTrialsAdvanced.aspx?protocolsearchid=5678174>). Significantly, some sporadic breast, ovarian, prostate, pancreatic and other tumours also possess HR defects due to mutation or epigenetic inactivation of HR components, suggesting that PARP inhibitors might be more broadly applicable. Furthermore, as other DDR pathways are frequently impaired in cancers, there may be additional situations where DDR inhibitors would display selective antitumour effects. Consistent with this idea, CHK1 inhibition reportedly

sensitizes p53-deficient cells to DNA-damaging agents more than p53-proficient cells⁹². The development of diagnostic procedures to identify DDR differences between cancer and normal cells therefore holds great promise for intelligent tailoring of DNA-damaging therapies and DDR-inhibitor therapies for the individual patient. Furthermore, as DDR activation is prevalent during oncogenesis, screening for DDR markers could enhance the reliability and sensitivity of cancer detection, and might allow effective detection of pre-malignant disease. In the longer term, it might be possible to develop drugs that enhance DDR events, thus reducing cancer incidence. In this regard, it is noteworthy that mice engineered to exhibit enhanced p53-dependent DNA-damage responses are less tumour prone than wild-type mice⁹³.

Ischaemia-reperfusion injury, inflammatory diseases and ageing.

Although DDR mechanisms generally protect against disease, their hyper-activation can contribute to pathology. A prime example of this is in ischaemia-reperfusion episodes associated with stroke and myocardial infarction, where PARP1 can become hyper-activated through DNA damage caused by re-oxygenation and nitric-oxide production. This depletes intracellular pools of nicotinamide adenine dinucleotide, resulting in impaired ATP production and cell death, often by necrosis. Notably, genetic inactivation or pharmacological inhibition of PARP1 in rodents provides considerable protection towards such cell death, consequently diminishing ensuing organ dysfunction. Similarly, animal models have shown PARP1 inhibition to protect against traumatic brain injury, endotoxic shock, tissue damage caused by chronic inflammation and drug-induced diabetes (for example see refs 94, 95). Thus, PARP1 inhibitors might find utility in treating such conditions in people. It is also noteworthy that p53 dysfunction is associated with inflammatory diseases and atherosclerosis^{96,97}, indicating that pharmacological modulation of p53, and its upstream activator ATM, might ameliorate such pathologies⁸⁵. With our growing realization that DNA damage and sub-optimal DDR events are associated with neurodegenerative disease and various other age-related degenerative pathologies, it is also tempting to speculate that DDR-modulatory drugs will one day be used to slow down or prevent such conditions; perhaps even certain aspects of the normal ageing process.

Viral, parasitic and other diseases. DDR proteins function in the life cycles of human parasites and pathogens, suggesting that DDR inhibitors could be used to treat their associated pathologies. For instance, the reliance of HIV on host-cell DDR factors suggests a potential for DDR inhibitors in AIDS therapy^{98,99}. Although such treatments would need to be evaluated for potential side effects, a possible advantage over conventional treatments that target the pathogen itself is that they would not be easily subject to evolution of resistance. Furthermore, antibacterial agents could be developed that target aspects of bacterial DDR mechanisms that are distinct from those of host cells.

Gene therapy. Correcting gene dysfunction is a long-sought-after treatment for many human maladies, including immune deficiencies, cystic fibrosis, muscular dystrophy and hereditary blindness. Although some success has been achieved, such approaches have been plagued by safety issues, largely arising through unwanted NHEJ-mediated integration of the introduced gene into tumour-suppressor loci. Although it is difficult to imagine such obstacles being surmounted in the near future, the development of methods to interfere with NHEJ or promote gene integration into desired loci (for example, see ref. 100) offers exciting prospects for gene-therapy optimization.

Future challenges

Great progress has been made towards understanding the DDR but much remains to be learned. One major future challenge is to understand in more detail how the activities of DDR proteins are controlled. Other challenges are to determine precisely how and why the DDR impacts on myriad cellular functions and how such complex programmes are orchestrated. Additional important issues to be

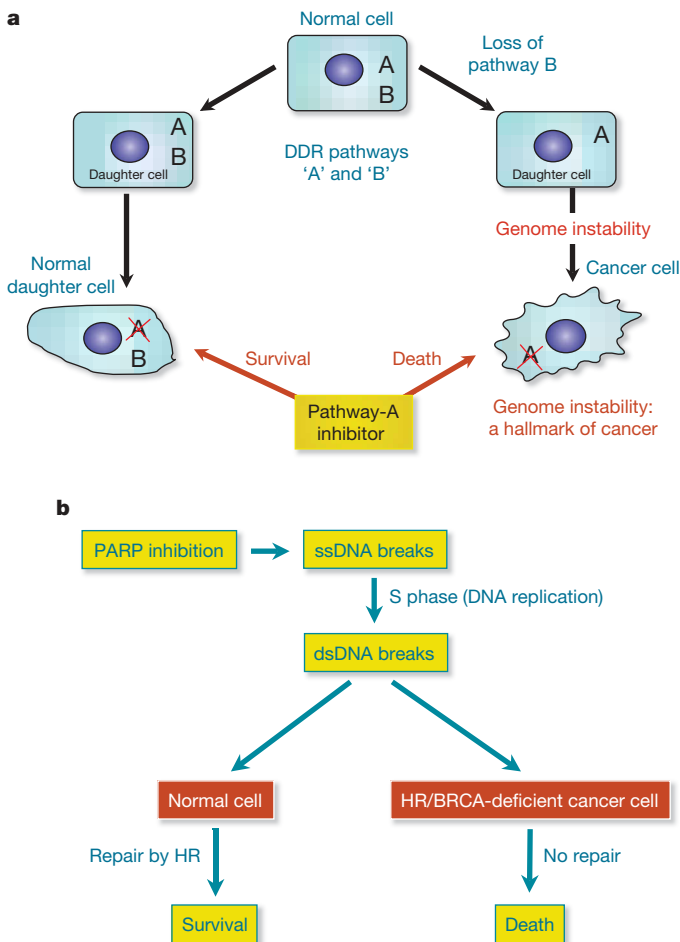


Figure 2 | Exploitation of DDR pathways to enhance therapeutic responses. **a**, Model: normal cells have two DDR pathways: A and B. If one pathway (B) is eliminated, genome instability results, which can foster the evolution of a cancer cell. Addition of an inhibitor targeting the second pathway (A) leads to cell death. Normal cells that still retain an active B pathway, however, survive. **b**, Treatment with a PARP inhibitor selectively kills HR/BRCA-deficient cells. PARP inhibition impairs the repair of SSBs, which are converted to DSBs in S phase. Such DSBs are effectively repaired by HR in non-cancerous cells but not in BRCA-deficient cancer cells. Adapted with permission from ref. 101.

addressed are how the DDR can be shaped and fine-tuned by other pathways and events, and how the same DDR stimulus can yield markedly different responses in different cells and tissues, including cancer cells and stem cells. Such knowledge will not only enhance our appreciation of DDR functions but will undoubtedly present exciting opportunities for better understanding and managing human health and disease.

1. Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* **65**, 127–134 (2000).
2. An excellent overview of the extent of endogenous DNA damage, the types of DNA lesions arising from cell autonomous sources, and the pathways that repair such lesions.
3. Valko, M., Rhodes, C. J., Moncol, J., Izakovic, M. & Mazur, M. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem. Biol. Interact.* **160**, 1–40 (2006).
4. Kawanishi, S., Hiraku, Y., Pinlaor, S. & Ma, N. Oxidative and nitrate DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis. *Biol. Chem.* **387**, 365–372 (2006).
5. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signalling, repair and the cancer connection. *Nature Genet.* **27**, 247–254 (2001).
6. Ward, J. F. DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *Prog. Nucleic Acid Res. Mol. Biol.* **35**, 95–125 (1988).
7. Doll, R. & Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl Cancer Inst.* **66**, 1191–1308 (1981).
8. Classical overview of epidemiological evidence for DNA-damaging environmental insults implicated as carcinogens, and suggestions for measures to prevent such tumours.
9. Wogan, G. N., Hecht, S. S., Felton, J. S., Conney, A. H. & Loeb, L. A. Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* **14**, 473–486 (2004).
10. Espinosa, E., Zamora, P., Feliu, J. & Gonzalez Baron, M. Classification of anticancer drugs—a new system based on therapeutic targets. *Cancer Treat. Rev.* **29**, 515–523 (2003).
11. Lebwohl, M., Ting, P. T. & Koo, J. Y. Psoriasis treatment: traditional therapy. *Ann. Rheum. Dis.* **64** (suppl. 2), 83–86 (2005).
12. Harper, J. W. & Elledge, S. J. The DNA damage response: ten years after. *Mol. Cell* **28**, 739–745 (2007).
13. Rouse, J. & Jackson, S. P. Interfaces between the detection, signalling, and repair of DNA damage. *Science* **297**, 547–551 (2002).
14. Harrison, J. C. & Haber, J. E. Surviving the Breakup: The DNA damage checkpoint. *Annu. Rev. Genet.* **40**, 209–235 (2006).
15. Jiricny, J. The multifaceted mismatch-repair system. *Nature Rev. Mol. Cell Biol.* **7**, 335–346 (2006).
16. David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950 (2007).
17. Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
18. A highly informative review of the links between DNA damage, DNA-repair pathways and their defects contributing to tumorigenesis.
19. Friedberg, E. C. *et al.* *DNA Repair and Mutagenesis* 2nd edn (ASM Press, 2006).
20. An excellent, comprehensive multi-author book covering essentially the entire field of DNA repair, from basic mechanisms in diverse organisms, to human diseases associated with defective DNA repair.
21. Loeb, L. A. & Monnat, R. J. Jr. DNA polymerases and human disease. *Nature Rev. Genet.* **9**, 594–604 (2008).
22. Lieber, M. R. The mechanism of human nonhomologous DNA end joining. *J. Biol. Chem.* **283**, 1–5 (2008).
23. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* **77**, 229–257 (2008).
24. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529–538 (2008).
25. Kennedy, R. D. & D'Andrea, A. D. The Fanconi Anemia/BRCA pathway: new faces in the crowd. *Genes Dev.* **19**, 2925–2940 (2005).
26. Cimprich, K. A. & Cortez, D. ATR: an essential regulator of genome integrity. *Nature Rev. Mol. Cell Biol.* **9**, 616–627 (2008).
27. Bartek, J. & Lukas, J. DNA damage checkpoints: from initiation to recovery or adaptation. *Curr. Opin. Cell Biol.* **19**, 238–245 (2007).
28. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).
29. Describes the key DDR kinases ATM, ATR and DNA-PK, provides an overview of their substrates, and outlines the cellular pathways affected by DNA-damage signalling.
30. Riley, T., Sontag, E., Chen, P. & Levine, A. Transcriptional control of human p53-regulated genes. *Nature Rev. Mol. Cell Biol.* **9**, 402–412 (2008).
31. Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
32. Huen, M. S. & Chen, J. The DNA damage response pathways: at the crossroad of protein modifications. *Cell Res.* **18**, 8–16 (2008).
33. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
34. Milestone report on the proteomic identification of ATM/ATR substrates and their assignment to various cellular functions, including RNA processing and other protein networks not previously recognized as DDR targets.
35. Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
36. Halazonetis, T. D., Gorgoulis, V. G. & Bartek, J. An oncogene-induced DNA damage model for cancer development. *Science* **319**, 1352–1355 (2008).
37. Misteli, T. & Soutoglou, E. The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Rev. Mol. Cell Biol.* **10**, 243–254 (2009).
38. Ziv, Y. *et al.* Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nature Cell Biol.* **8**, 870–876 (2006).
39. Xiao, A. *et al.* WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature* **457**, 57–62 (2009).
40. Cook, P. J. *et al.* Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature* **458**, 591–596 (2009).
41. Bassing, C. H. & Alt, F. W. The cellular response to general and programmed DNA double-strand breaks. *DNA Repair* **3**, 781–796 (2004).
42. Schlissel, M. S., Kaffer, C. R. & Curry, J. D. Leukemia and lymphoma: a cost of doing business for adaptive immunity. *Genes Dev.* **20**, 1539–1544 (2006).
43. Richardson, C., Horikoshi, N. & Pandita, T. K. The role of the DNA double-strand break response network in meiosis. *DNA Repair* **3**, 1149–1164 (2004).
44. Verdun, R. E. & Karlseder, J. Replication and protection of telomeres. *Nature* **447**, 924–931 (2007).
45. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
46. d'Adda di Fagagna, F., Teo, S. H. & Jackson, S. P. Functional links between telomeres and proteins of the DNA-damage response. *Genes Dev.* **18**, 1781–1799 (2004).
47. References 39 and 40 illustrate the intimate links between the telomere maintenance and DDR machineries.
48. Longhese, M. P. DNA damage response at functional and dysfunctional telomeres. *Genes Dev.* **22**, 125–140 (2008).
49. Sedelnikova, O. A. *et al.* Senescing human cells and ageing mice accumulate DNA lesions with unreparable double-strand breaks. *Nature Cell Biol.* **6**, 168–170 (2004).
50. Niedernhofer, L. J. *et al.* A new progeroid syndrome reveals that genotoxic stress suppresses the somatotrophic axis. *Nature* **444**, 1038–1043 (2006).
51. Reports a powerful mouse model of multifaceted premature ageing, based on engineered deficiency in the Xpf gene involved in transcription-coupled NER.
52. Jayapalan, J. C. & Sedivy, J. M. Cellular senescence and organismal aging. *Mech. Ageing Dev.* **129**, 467–474 (2008).
53. Nospikel, T. P., Hyka-Nospikel, N. & Hanawalt, P. C. Transcription domain-associated repair in human cells. *Mol. Cell Biol.* **26**, 8722–8730 (2006).
54. Orii, K. E., Lee, Y., Kondo, N. & McKinnon, P. J. Selective utilization of nonhomologous end-joining and homologous recombination DNA repair pathways during nervous system development. *Proc. Natl Acad. Sci. USA* **103**, 10017–10022 (2006).
55. Park, Y. & Gerson, S. L. DNA repair defects in stem cell function and aging. *Annu. Rev. Med.* **56**, 495–508 (2005).
56. Nijnik, A. *et al.* DNA repair is limiting for haematopoietic stem cells during ageing. *Nature* **447**, 686–690 (2007).
57. Rossi, D. J. *et al.* Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447**, 725–729 (2007).
58. Collis, S. J. & Boulton, S. J. Emerging links between the biological clock and the DNA damage response. *Chromosoma* **116**, 331–339 (2007).
59. Collis, S. J. *et al.* HCLK2 is essential for the mammalian S-phase checkpoint and impacts on Chk1 stability. *Nature Cell Biol.* **9**, 391–401 (2007).
60. Kang, T. H., Reardon, J. T., Kemp, M. & Sancar, A. Circadian oscillation of nucleotide excision repair in mammalian brain. *Proc. Natl Acad. Sci. USA* **106**, 2864–2867 (2009).
61. Writing Committee of the Second World Health Organization Consultation on Clinical Aspects of Human Infection with Avian Influenza A (H5N1) Virus.. Update on avian influenza A (H5N1) virus infection in humans. *N. Engl. J. Med.* **358**, 261–273 (2008).
62. McCulloch, R. & Barry, J. D. A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation. *Genes Dev.* **13**, 2875–2888 (1999).
63. Lilley, C. E., Schwartz, R. A. & Weitzman, M. D. Using or abusing: viruses and the cellular DNA damage response. *Trends Microbiol.* **15**, 119–126 (2007).
64. Narisawa-Saito, M. & Kiyono, T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. *Cancer Sci.* **98**, 1505–1511 (2007).
65. Muylaert, I. & Elias, P. Knock-down of DNA ligase IV/ XRCC4 by RNAi inhibits herpes simplex virus type I DNA replication. *J. Biol. Chem.* **282**, 10865–10872 (2007).
66. Li, H. *et al.* Chk2 is required for HSV-1 ICPO-mediated G2/M arrest and enhancement of virus growth. *Virology* **375**, 13–23 (2008).
67. Smith, J. & Daniel, R. Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. *ACS Chem. Biol.* **1**, 217–226 (2006).
68. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
69. A comprehensive overview of cancer-predisposing mutations and advances in cancer genetics.

61. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
62. Maser, R. S. & DePinho, R. A. Connecting chromosomes, crisis, and cancer. *Science* **297**, 565–569 (2002).
63. Bristow, R. G. & Hill, R. P. Hypoxia and metabolism. Hypoxia, DNA repair and genetic instability. *Nature Rev. Cancer* **8**, 180–192 (2008).
64. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
65. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
- References 64 and 65 provide evidence for activation of the DDR machinery in early human oncogenic lesions and models of oncogenic transformation, and propose that the DNA-damage checkpoint activated by oncogene-evoked replication stress and DNA breakage provides an inducible barrier against tumour progression.**
66. Rass, U., Ahel, I. & West, S. C. Defective DNA repair and neurodegenerative disease. *Cell* **130**, 991–1004 (2007).
67. Kulkarni, A. & Wilson, D. M. III. The involvement of DNA-damage and -repair defects in neurological dysfunction. *Am. J. Hum. Genet.* **82**, 539–566 (2008).
68. Weissman, L., de Souza-Pinto, N. C., Stevnsner, T. & Bohr, V. A. DNA repair, mitochondria, and neurodegeneration. *Neuroscience* **145**, 1318–1329 (2007).
69. Caldecott, K. W. Single-strand break repair and genetic disease. *Nature Rev. Genet.* **9**, 619–631 (2008).
70. Ljungman, M. & Lane, D. P. Transcription—guarding the genome by sensing DNA damage. *Nature Rev. Cancer* **4**, 727–737 (2004).
71. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
72. Kovtun, I. V. & McMurray, C. T. Features of trinucleotide repeat instability *in vivo*. *Cell Res.* **18**, 198–213 (2008).
73. Yang, J. L., Weissman, L., Bohr, V. A. & Mattson, M. P. Mitochondrial DNA damage and repair in neurodegenerative disorders. *DNA Repair* **7**, 1110–1120 (2008).
74. Matzuk, M. M. & Lamb, D. J. The biology of infertility: research advances and clinical challenges. *Nature Med.* **14**, 1197–1213 (2008).
75. Bartkova, J., Rajpert-De Meyts, E., Skakkebaek, N. E., Lukas, J. & Bartek, J. DNA damage response in human testes and testicular germ cell tumours: biology and implications for therapy. *Int. J. Androl.* **30**, 282–291 (2007).
76. Schumacher, B., Garinis, G. A. & Hoeijmakers, J. H. Age to survive: DNA damage and aging. *Trends Genet.* **24**, 77–85 (2008).
- A thought-provoking review of the evidence for causative links between DNA-damage accumulation and organismal ageing, which proposes the concept of a survival response that allows the organism's resources to be shifted from emphasis on growth, to survival of DNA damage and other stresses.**
77. Herbig, U., Ferreira, M., Condel, L., Carey, D. & Sedivy, J. M. Cellular senescence in aging primates. *Science* **311**, 1257 (2006).
78. Sharpless, N. E. & DePinho, R. A. How stem cells age and why this makes us grow old. *Nature Rev. Mol. Cell Biol.* **8**, 703–713 (2007).
79. Navarro, S. *et al.* Hematopoietic dysfunction in a mouse model for Fanconi anemia group D1. *Mol. Ther.* **14**, 525–535 (2006).
80. Reese, J. S., Liu, L. & Gerson, S. L. Repopulating defect of mismatch repair-deficient hematopoietic stem cells. *Blood* **102**, 1626–1633 (2003).
81. Mocanu, M. M. & Yellon, D. M. p53 down-regulation: a new molecular mechanism involved in ischaemic preconditioning. *FEBS Lett.* **555**, 302–306 (2003).
82. Vousden, K. H. & Lane, D. P. p53 in health and disease. *Nature Rev. Mol. Cell Biol.* **8**, 275–283 (2007).
83. Mercer, J., Mahmoudi, M. & Bennett, M. DNA damage, p53, apoptosis and vascular disease. *Mutat. Res.* **621**, 75–86 (2007).
84. Schneider, J. G. *et al.* ATM-dependent suppression of stress signaling reduces vascular disease in metabolic syndrome. *Cell Metab.* **4**, 377–389 (2006).
85. Kastan, M. B. DNA damage responses: mechanisms and roles in human disease. 2007 G.H.A. Clowes Memorial Award Lecture. *Mol. Cancer Res.* **6**, 517–524 (2008).
86. Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756–760 (2006).
87. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nature Rev. Cancer* **8**, 193–204 (2008).
88. Martin, S. A., Lord, C. J. & Ashworth, A. DNA repair deficiency as a therapeutic target in cancer. *Curr. Opin. Genet. Dev.* **18**, 80–86 (2008).
89. Jiang, H. *et al.* The combined status of ATM and p53 link tumor development with therapeutic response. *Genes Dev.* **23**, 1895–1909 (2009).
90. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
91. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
- References 90 and 91 document the potential of personalized cancer treatment, based on the exceptional sensitivity of tumour cells defective in BRCA1/BRCA2-dependent HR towards small molecule inhibitors of PARP1; these studies support the principle of synthetic-lethal relationships between complementary DDR pathways.**
92. Chen, Z. *et al.* Selective Chk1 inhibitors differentially sensitize p53-deficient cancer cells to cancer therapeutics. *Int. J. Cancer* **119**, 2784–2794 (2006).
93. García-Cao, I. *et al.* 'Super p53' mice exhibit enhanced DNA damage response, are tumor resistant and age normally. *EMBO J.* **21**, 6225–6235 (2002).
94. Pacher, P. & Szabo, C. Role of poly(ADP-ribose) polymerase 1 (PARP-1) in cardiovascular diseases: the therapeutic potential of PARP inhibitors. *Cardiovasc. Drug Rev.* **25**, 235–260 (2007).
95. Moroni, F. Poly(ADP-ribose)polymerase 1 (PARP-1) and postischemic brain damage. *Curr. Opin. Pharmacol.* **8**, 96–103 (2008).
96. Guevara, N. V., Kim, H. S., Antonova, E. I. & Chan, L. The absence of p53 accelerates atherosclerosis by increasing cell proliferation *in vivo*. *Nature Med.* **5**, 335–339 (1999).
97. Andreassi, M. G. DNA damage, vascular senescence and atherosclerosis. *J. Mol. Med.* **86**, 1033–1043 (2008).
98. Lau, A. *et al.* Suppression of HIV-1 infection by a small molecule inhibitor of the ATM kinase. *Nature Cell Biol.* **7**, 493–500 (2005).
99. Smith, J. A. *et al.* Evidence that the Nijmegen breakage syndrome protein, an early sensor of double-strand DNA breaks (DSB), is involved in HIV-1 post-integration repair by recruiting the ataxia telangiectasia-mutated kinase in a process similar to, but distinct from, cellular DSB repair. *Virology* **311**, 11 (2008).
100. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl Acad. Sci. USA* **104**, 3055–3060 (2007).
101. Jackson, S. P. The DNA-damage response: new molecular insights and new approaches to cancer therapy. *Biochem. Soc. Trans.* **37**, 483–494 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Polo and P. Huertas for advice, and K. Dry for expert help with the text and figures. The S.P.J. laboratory is supported by grants from Cancer Research UK, the European Commission (projects GENICA and DNA Repair), the Wellcome Trust and the Biotechnology and Biological Sciences Research Council. The J.B. laboratory is supported by grants from the Danish Cancer Society, the Danish National Research Foundation and the European Commission (projects GENICA, Active p53, TRIEME and DNA Repair).

Author Contributions S.P.J. and J.B. conceived of and wrote all aspects of this article.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to S.P.J. (s.jackson@gurdon.cam.ac.uk).

A role for a neo-sex chromosome in stickleback speciation

Jun Kitano^{1†}, Joseph A. Ross^{1,2†}, Seiichi Mori³, Manabu Kume⁴, Felicity C. Jones⁵, Yingguang F. Chan⁵, Devin M. Absher^{6†}, Jane Grimwood^{6†}, Jeremy Schmutz^{6†}, Richard M. Myers^{6†}, David M. Kingsley⁵ & Catherine L. Peichel¹

Sexual antagonism, or conflict between the sexes, has been proposed as a driving force in both sex-chromosome turnover and speciation. Although closely related species often have different sex-chromosome systems, it is unknown whether sex-chromosome turnover contributes to the evolution of reproductive isolation between species. Here we show that a newly evolved sex chromosome contains genes that contribute to speciation in threespine stickleback fish (*Gasterosteus aculeatus*). We first identified a neo-sex chromosome system found only in one member of a sympatric species pair in Japan. We then performed genetic linkage mapping of male-specific traits important for reproductive isolation between the Japanese species pair. The neo-X chromosome contains loci for male courtship display traits that contribute to behavioural isolation, whereas the ancestral X chromosome contains loci for both behavioural isolation and hybrid male sterility. Our work not only provides strong evidence for a large X-effect on reproductive isolation in a vertebrate system, but also provides direct evidence that a young neo-X chromosome contributes to reproductive isolation between closely related species. Our data indicate that sex-chromosome turnover might have a greater role in speciation than was previously appreciated.

Sexually antagonistic selection has been proposed as a major driving force in the evolution of sex chromosomes. Specifically, natural selection is expected to favour linkage between genes with sexually antagonistic effects (that is, beneficial in one sex and detrimental in the other) and the sex-determination locus, resulting in a reduction of recombination between sex chromosomes¹. In an XY sex-chromosome system, a reduction in recombination ultimately leads to the degeneration of the Y chromosome, thereby exposing alleles on the hemizygous X chromosome to selection in males². Thus, male-beneficial alleles, manifested as sexually dimorphic and/or sexually selected traits, are predicted to accumulate on the X chromosome^{3,4}. When these male-beneficial traits or alleles are important for reproductive isolation between species, the X chromosome is also predicted to have a key role in speciation. A disproportionately large effect of the X chromosome has been demonstrated for hybrid male sterility^{5,6}, although the data supporting a large X-effect for other isolating barriers has been less consistent⁶.

Sexually antagonistic selection is also predicted to drive the divergence of sex-chromosome systems between closely related species^{7,8}. Sex-chromosome turnover has been observed across many taxa, but it is particularly striking in fishes^{9,10}. Many independent groups of fishes show evidence for the rapid evolution of sex chromosomes through several different mechanisms, including the transposition of an existing male-determination locus to an autosome¹¹, the evolution of a new male-determination locus on an autosome¹², and fusions between an autosome and an existing Y chromosome^{13,14}. It has been suggested that the rapid turnover of sex chromosomes driven by sexual conflict might also have a role in the high speciation rates seen in some groups of fishes^{15–17}. However, a direct role for sex-chromosome turnover in speciation has not been empirically investigated in any system.

Here we demonstrate that there is a newly formed sex-chromosome system in a threespine stickleback population found in Japan. This Japan Sea form diverged from the Pacific Ocean threespine stickleback during periods of geographical isolation between the Sea of Japan and the Pacific Ocean about 1.5–2 million years ago^{18,19}. Because we can cross the derived Japan Sea form to the ancestral Pacific Ocean form, we have been able to take advantage of the genetic tools available for the threespine stickleback²⁰ to investigate whether the evolution of a neo-sex chromosome has contributed to the evolution of male traits that act as isolating barriers between the derived Japan Sea and the ancestral Pacific Ocean populations.

Japan Sea stickleback neo-sex chromosome

In all threespine stickleback populations examined previously, including the Pacific Ocean form, linkage group 19 (LG19) is the sex chromosome and LG9 is an autosome^{14,21,22}. However, in the course of making a linkage map from a Japan Sea cross, we noticed that several LG9 markers co-segregated with LG19 markers previously found to be tightly linked to the sex-determination locus in a region of reduced recombination and rearrangements on the Y chromosome^{21,22}. This association was observed when male meioses, but not female meioses, were analysed (Fig. 1a). These data indicated that one copy of LG9 might be fused to one copy of LG19 (the Pacific Ocean Y chromosome), forming a neo-Y chromosome in Japan Sea sticklebacks. Therefore, we performed fluorescence *in situ* hybridization (FISH) with LG19 and LG9 probes on metaphase chromosome spreads from two different populations of the Japan Sea form. We found that the Japan Sea males ($n = 4$) from two different populations have an odd number of chromosomes ($2n = 41$), with one large

¹Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. ²Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington 98195, USA. ³Biological Laboratory, Gifu-keizai University, Ogaki, Gifu 503-8550, Japan. ⁴Aqua Restoration Research Center, Public Works Research Institute, Kakamigahara, Gifu 501-6021, Japan. ⁵Department of Developmental Biology and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. ⁶Department of Genetics and Stanford Human Genome Center, Stanford University, Stanford, California 94304, USA. [†]Present addresses: Department of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan (J.K.); Department of Biology, University of Maryland, College Park, Maryland 20742, USA (J.A.R.); HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA (D.M.A., J.G., J.S. and R.M.M.)

unpaired chromosome that hybridized to the LG19 and LG9 probes, providing evidence for the LG9–Y chromosome fusion (Fig. 1b and Supplementary Fig. 1). In contrast, the Japan Sea females ($n = 4$) from both populations have an even number of chromosomes ($2n = 42$), and the LG9 and LG19 probes hybridize to separate chromosome pairs (Fig. 1b and Supplementary Fig. 1). Because the fused copy of LG9 segregates with the Y chromosome in Japan Sea males, the other copy segregates as an X chromosome; this neo-sex chromosome system is defined as an X_1X_2Y system¹³, in which X_1 is the ancestral X chromosome (LG19) and X_2 is the neo-X chromosome (LG9). Thus, the Japan Sea form has a unique neo-sex chromosome system, which has probably evolved within the past 1.5–2 million years of isolation between the Pacific Ocean and Japan Sea sticklebacks^{18,19}.

Components of reproductive isolation

We have the opportunity to test the role of this neo-sex chromosome in reproductive isolation between the Japan Sea and the Pacific Ocean forms because both are anadromous and migrate into Lake Akkeshi and the Bikanbeushi marsh during the breeding season. In this location, they co-occur, along with hybrid adults and juveniles, within a hybrid zone in the Bikanbeushi River (Mid2 in Fig. 2, and Supplementary Fig. 2). We found that temporal isolation, behavioural isolation and hybrid male sterility contribute to reproductive isolation between the forms in the hybrid zone (Supplementary Fig. 3 and Supplementary Discussion). We previously demonstrated that behavioural isolation and hybrid male sterility are asymmetric, but act in complementary directions¹⁹. Pacific Ocean females mate exclusively with Pacific Ocean males, whereas the Japan Sea females mate

with both types of males in laboratory mate-choice trials¹⁹. Although Japan Sea females do mate with Pacific Ocean males, hybrid males resulting from this cross have severely reduced fertility, whereas reciprocal hybrid males and all hybrid females are fertile¹⁹.

To test whether reproductive isolation is linked to sex-chromosome divergence, we first investigated which male mating traits contribute to asymmetric behavioural isolation. First, we analysed the relationship between final female choice and differences in male body size and found that both types of females tend to choose larger males (Fig. 3a; Pacific Ocean females, $n = 30$, coefficient estimate = 0.254, $Z = 2.276$, $P = 0.0228$, logistic regression; Japan Sea females, $n = 29$, coefficient estimate = 0.122, $Z = 2.349$, $P = 0.0188$, logistic regression). Because Japan Sea males (standard length = 63.44 ± 0.37 mm (mean \pm s.e.m.), $n = 59$) are smaller than Pacific Ocean males (standard length = 76.17 ± 0.47 mm, $n = 45$; analysis of variance (ANOVA), $F_{1,102} = 475$, $P < 10^{-15}$), divergence in body size is one of the factors that contributes to asymmetric behavioural isolation. However, even in the absence of body size divergence, Pacific Ocean females still have a 93.5% probability (95% confidence interval = 0.624–0.992) of choosing a conspecific male (Fig. 3a), suggesting that other factors are involved in Pacific Ocean female choice.

We found that a difference in male dorsal pricking behaviour also contributes to asymmetric behavioural isolation. Dorsal pricking is a component of male mating behaviour specific to threespine sticklebacks in which the male raises his dorsal spines and pricks the female. This behaviour might help the male assess the female, provide tactile stimulation to induce female spawning, or act as a way for the male to display his dorsal spines to the female^{23,24}. In Japan Sea males, the

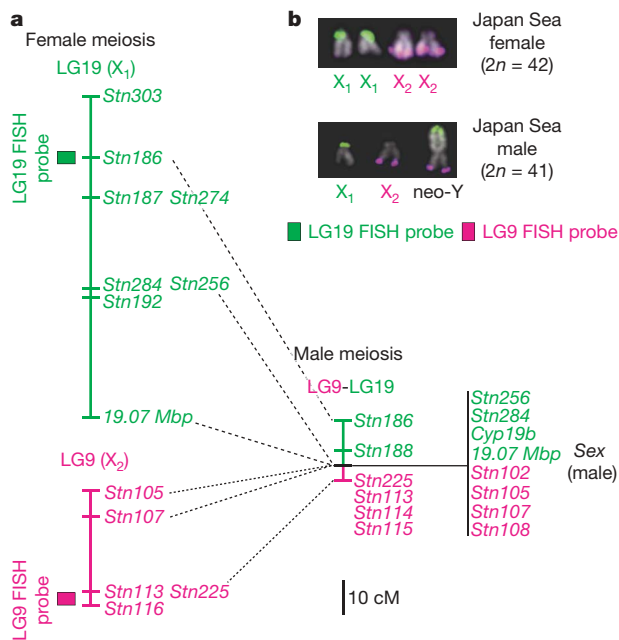


Figure 1 | Genetic and cytogenetic evidence of a fusion between one copy of LG9 and the Y chromosome in Japan Sea males. **a**, A cross between Japan Sea individuals was used to create a linkage map. The female meiotic maps (X–X recombination) of LG19 (X_1) and LG9 (X_2) are to the left, and the male meiotic map (X–Y recombination) of LG9/19 (neo-Y) is to the right. LG19 (X_1) is in green and LG9 (X_2) is in magenta. Genetic distances between the markers are drawn to scale. Scale bar, 10 centimorgans (cM). Several LG9 and LG19 markers do not recombine with each other or with the sex-determination locus (Sex) in males. The coloured bars to the left of the map indicate the relative position of the FISH probes. **b**, The X_1 and X_2 chromosomes from a representative Akkeshi Japan Sea female metaphase spread and the X_1 , X_2 and neo-Y chromosomes from a representative Akkeshi Japan Sea male metaphase spread are shown. A LG9 BAC (magenta) and a LG19 BAC (green) were used as probes for FISH.

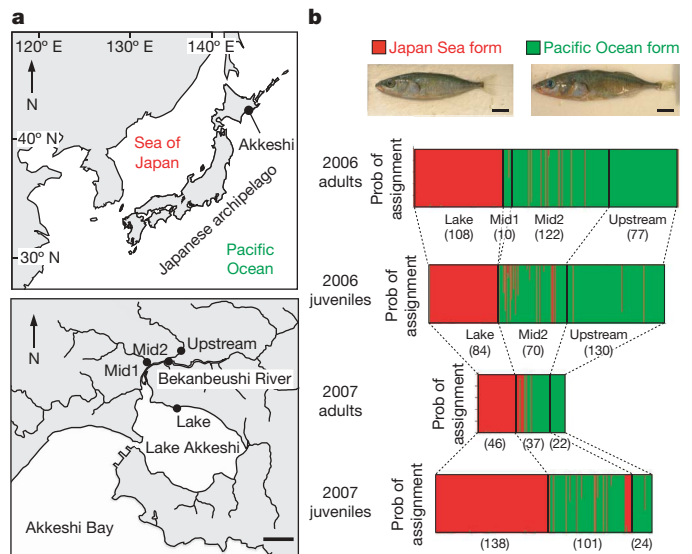


Figure 2 | Distribution of the Japanese threespine stickleback species pair in a region of sympatry. **a**, The top panel shows a map of the Japanese archipelago, indicating the location of Akkeshi in eastern Hokkaido, Japan. The bottom panel shows a map of our four collection sites (Lake, Mid1, Mid2 and Upstream) in Lake Akkeshi and the Bikanbeushi marsh. Scale bar, 2 km. **b**, The top panels show representative photos of a Japan Sea male and a Pacific Ocean male. Scale bar, 10 mm. Genetic analysis of fish collected in 2006 ($n = 601$) and 2007 ($n = 368$) was performed using STRUCTURE. There are two genetic clusters in Akkeshi (Supplementary Fig. 2), with the Japan Sea cluster shown in red, and the Pacific Ocean cluster shown in green. In the STRUCTURE plots, each individual is indicated by a single vertical line. The probability of assignment to the Japan Sea cluster or the Pacific Ocean cluster is indicated by the extent of the coloured bar. Fish collected from the same location are grouped together, with locations separated by thick black lines. The number of individuals per location is indicated in parenthesis. The Japan Sea form is mostly found in the lake, whereas the Pacific Ocean form is mostly found in the upstream region of the Bikanbeushi River. The midstream region (Mid2) is a hybrid zone that contains both parental forms and hybrids.

dorsal pricking display is greatly exaggerated, and the male pushes the female upwards during dorsal pricking (Fig. 3b)¹⁹. Furthermore, dorsal spine length is sexually dimorphic (males have longer spines than females) in the Japan Sea form²⁵. In contrast, in the Pacific Ocean form, the dorsal pricking display is weak (Fig. 3b), and dorsal spine length is not sexually dimorphic^{19,25}. We found that the Pacific Ocean females frequently escaped and did not resume mating after they encountered the aggressive dorsal pricking of Japan Sea males, whereas the Japan Sea females did not escape from males after dorsal pricking (Fig. 3b).

Genetic mapping of isolating barriers

To investigate the chromosomal locations of the isolating barriers between the two forms, we backcrossed F₁ hybrid (Japan Sea female × Pacific Ocean male) females to Pacific Ocean males and conducted quantitative trait locus (QTL) mapping of male dorsal pricking, male dorsal spine length, male body size, and hybrid male sterility. Individuals were genotyped with 90 single nucleotide polymorphism (SNP) markers and 13 other microsatellite markers that together span the stickleback genome. Hybrid male sterility and male body size mapped to LG19 (Fig. 4, Supplementary Fig. 4 and Supplementary Table 1), which is the ancestral X chromosome shared by the Japan Sea form and Pacific Ocean form. Dorsal pricking and first dorsal spine length mapped to distinct locations on LG9 (Fig. 4, Supplementary

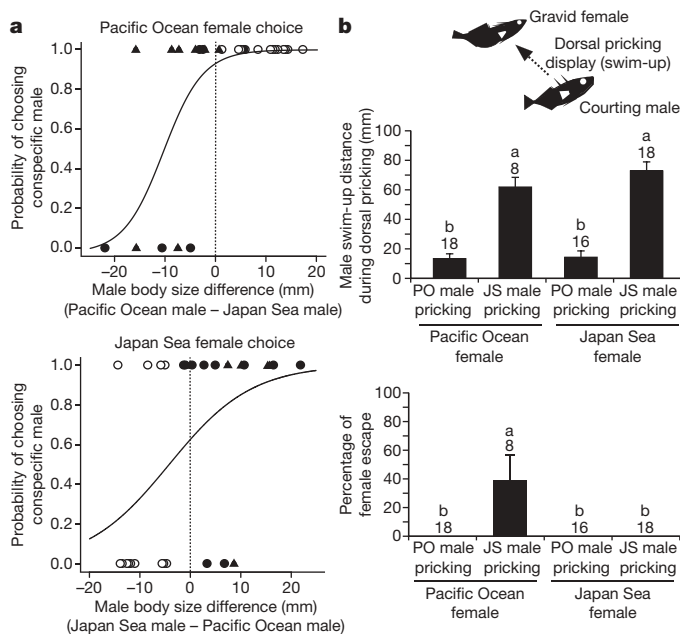


Figure 3 | Behavioural isolation results from divergence in male body size and male dorsal pricking behaviour. **a**, The top panel indicates the preferences of Pacific Ocean females ($n = 30$), and the bottom panel indicates the preferences of Japan Sea females ($n = 29$) as a function of male body size divergence. The horizontal axes indicate the body size difference in mm between males (conspecific male standard length minus heterospecific male standard length). Each symbol indicates a mate choice trial, in which 1 indicates the female chose a conspecific male, and 0 indicates the female chose a heterospecific male. Trials were conducted with laboratory-reared Pacific Ocean males (solid circles), resident freshwater Pacific males (solid triangles), or without size manipulation (open circles). The logistic regression curves indicate the probability of a female choosing a conspecific male for a given difference in male body size. **b**, The top panel shows the distance a Pacific Ocean (PO) or Japan Sea (JS) male moved upwards during dorsal pricking with Pacific Ocean or Japan Sea females. The bottom panel shows the percentage of trials in which a female escaped after dorsal pricking by either a Pacific Ocean or a Japan Sea male. The sample size for each mating pair is shown above each column. Lower case letters above the bars represent pairs that are significantly different from each other (pairwise Mann–Whitney U -test, $P < 0.05$ after Bonferroni correction).

Fig. 4 and Supplementary Table 1), which is the neo-X chromosome in the Japan Sea form. A genome-wide scan for epistatic interactions identified a significant conspecific interaction between loci on LG19 for hybrid male sterility (log likelihood ratio of linkage (LOD) comparing the full model with interaction to the additive model = 5.08; genome-wide significance threshold = 3.99 ($\alpha = 0.05$); Supplementary Fig. 5). No significant epistatic interactions were found for any other traits or loci examined.

Our data demonstrate that loci important for both prezygotic and postzygotic isolation map to the X chromosomes in this natural vertebrate system. This large X-effect is unlikely to result from an over-representation of these chromosomes in the stickleback genome, as LG9 and LG19 comprise just 9.0% of the stickleback genome (20.2 megabases (Mb) each of 446.6 Mb total—4.5% for each linkage group in the stickleback genome assembly; BROAD S1, Feb 2006). Because mapping in a backcross probably overestimates the effects of the hemizygous X chromosome²⁶, we also performed QTL mapping of the same traits in an independent F₂ intercross to ensure that the observed large X-effect was not simply the result of our backcross mapping strategy. We still detect QTL for dorsal pricking and first dorsal spine length on LG9 and QTL for male sterility and body size on LG19, with the addition of a single QTL for testis size on LG1 (Supplementary Fig. 6 and Supplementary Table 2).

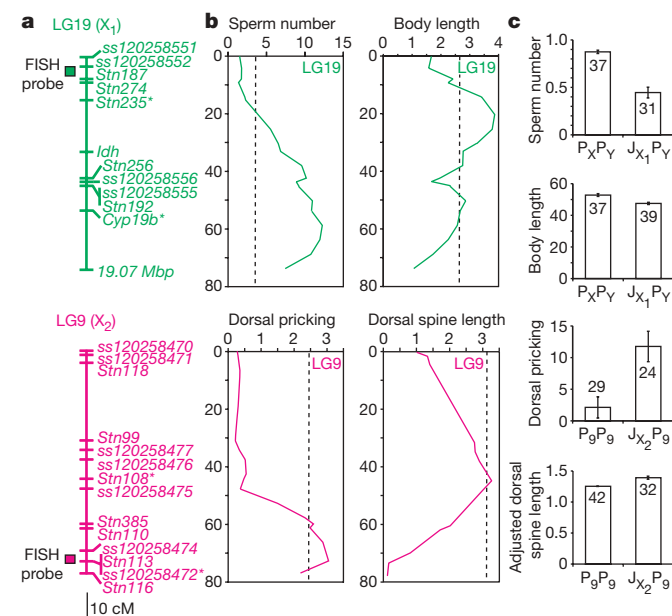


Figure 4 | Genetic mapping of isolating barriers. **a**, Genetic linkage maps of LG19 (X₁; green) and LG9 (X₂; magenta) in the backcross. The locations of FISH probes are indicated to the left of the map, whereas the names of genetic markers are indicated to the right. The asterisk indicates the marker closest to the QTL peak. **b**, For each QTL, the LOD score is indicated across the top and is plotted relative to the positions of the markers indicated in **a**, with distance in cM indicated. The hybrid male sterility QTL is represented by sperm number, the body size QTL is represented by body length, and the dorsal pricking QTL is represented by mean dorsal pricking. The dorsal spine length QTL is represented by first dorsal spine length, which was analysed with body length as an interacting covariate. Dashed lines indicate the genome-wide significance thresholds determined by permutation tests ($\alpha = 0.05$). **c**, For each trait, the phenotypic values (mean \pm s.e.m.) are indicated for genotypes at the marker closest to the QTL peak; that is, *Cyp19b* for sperm number, *Stn235* for body length, SNP ss120258472 for mean dorsal pricking, and *Stn108* for dorsal spine length. Sample sizes for each genotypic class are shown in the graph. These were the only genomic regions with significant phenotype-genotype associations after Bonferroni correction ($P < 0.0001$), detected either by a Kruskal–Wallis test (sperm number, body length, and dorsal pricking) or by analysis of covariance (ANCOVA) (dorsal spine length).

Discussion

Mapping of hybrid male sterility to the ancestral X chromosome (LG19) is consistent with previous studies on hybrid male sterility in other systems^{5,6}, suggesting that the large X-effect on hybrid male sterility is common across diverse taxa. Although the reasons for the large X-effect on hybrid sterility are debated⁵, in some cases it results from genetic conflict in the form of sex ratio meiotic drive^{27,28}. However, we find no evidence for sex ratio distortion in the weakly fertile F₁ hybrid males resulting from a cross between Japan Sea females and Pacific Ocean males (Supplementary Table 3). We do, however, find evidence for conspecific epistasis between X-linked loci, which is commonly observed for hybrid male sterility in *Drosophila*^{29,30}. Despite this finding of a common large X-effect across widely separated taxa, we found no evidence for any effect of the neo-X chromosome on hybrid male sterility. It may be that the relative age and/or levels of degeneration of sex chromosomes are important factors in determining whether the X chromosome contributes to the evolution of hybrid male sterility.

Unlike male sterility, male courtship display traits conferring behavioural isolation between the Japan Sea and Pacific Ocean forms map to both the ancestral X chromosome and the neo-X chromosome. Interestingly, male body size is sexually dimorphic in both the Japan Sea and Pacific Ocean forms²⁵ and maps to the ancestral X chromosome. In contrast, first dorsal spine length is only sexually dimorphic in the Japan Sea form²⁵, and the dorsal pricking display is exaggerated in the Japan Sea males; these traits both map to the neo-X chromosome. Thus, these display traits might have evolved as a result of differential fitness effects between males and females specifically in the Japan Sea lineage. Although our cross design did not allow us to directly test whether these traits mapped to the neo-Y chromosome as predicted by theory^{7,8}, it is possible that selection for linkage between male beneficial traits, such as dorsal pricking and dorsal spine length, and the sex-determination locus actually promoted the spread of the fusion between LG9 and the Y chromosome in the Japan Sea population⁷. Alternatively, these male beneficial traits may have accumulated on the neo-X chromosome³ after the formation of the neo-Y chromosome in the last 1.5–2 million years. In either case, our data provide direct empirical evidence linking sex chromosome turnover and reproductive isolation between closely related species.

The turnover of sex chromosomes between species is common in many groups of animals, where closely related species often differ in sex-determination and sex-chromosome systems^{9,10}. Although it has been suggested that sex-chromosome turnover might drive rapid speciation in cichlid fishes in which a female colour trait is linked to an invading ZW sex-chromosome system^{15–17}, this hypothesis has not been directly tested. Given the potential role of sexual antagonism in driving both sex-chromosome evolution^{1,3,7,8} and speciation^{31,32}, we suggest that sex-chromosome divergence between closely related species should be given further consideration as an important mechanism contributing to the evolution of reproductive isolation.

METHODS SUMMARY

Threespine sticklebacks of the Pacific Ocean and the Japan Sea form were collected with seine nets and minnow traps in Lake Akkeshi and the Beukanbeushi River in May–July of 2003–2008. Japan Sea fish were also collected from an allopatric site (Cape of Benkei) on the west coast of Hokkaido in 2008. For cytogenetic and behavioural studies, live fish were transported to the Fred Hutchinson Cancer Research Center; all experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC number 1575). Cytogenetics and FISH were performed on Japan Sea males ($n = 2$) and a female ($n = 1$) from Akkeshi, as well as Japan Sea males ($n = 2$) and females ($n = 3$) from the allopatric site, using fluorescently labelled bacterial artificial chromosome (BAC) clones as previously described²². For population genetic analysis, 969 fish were genotyped with 12 neutral microsatellite markers (Supplementary Table 4). Mate-choice experiments were conducted as previously described¹⁹. For QTL mapping, a Japan Sea female and a Pacific Ocean

male were crossed to obtain an F₁ hybrid family ($J_1 \times P_1$). These F₁ females were crossed with males resulting from a cross between a Pacific Ocean female and another Pacific Ocean male ($P_2 \times P_3$) to generate backcross progeny. At maturity, 76 backcross males were phenotyped for traits related to body size, fertility and dorsal pricking behaviour. These males were genotyped with LG9 and LG19 microsatellites, as well as a panel of SNP markers distributed across the stickleback genome (Supplementary Table 5). All DNA isolation and microsatellite genotyping were conducted as previously described¹⁹; SNP genotyping was performed using Illumina Golden Gate arrays. The genotypes of 90 SNPs and 14 microsatellites were used to create a linkage map in JoinMap 3.0 (ref. 33), and QTL analyses were performed in MapQTL 4.0 (ref. 34) and R/qtl³⁵.

Received 4 May; accepted 20 August 2009.

Published online 27 September 2009.

1. Rice, W. R. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**, 911–914 (1987).
2. Rice, W. R. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167 (1987).
3. Rice, W. R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742 (1984).
4. Charlesworth, B., Coyne, J. A. & Barton, N. H. The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).
5. Presgraves, D. C. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* **24**, 336–343 (2008).
6. Qvarnström, A. & Bailey, R. I. Speciation through evolution of sex-linked genes. *Heredity* **102**, 4–15 (2009).
7. Charlesworth, D. & Charlesworth, B. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214 (1980).
8. van Doorn, G. S. & Kirkpatrick, M. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**, 909–912 (2007).
9. Ezaz, T., Stiglec, R., Veyrunes, F. & Graves, J. A. M. Relationships between vertebrate ZW and XY sex chromosome systems. *Curr. Biol.* **16**, R736–R743 (2006).
10. Mank, J. E., Promislow, D. E. L. & Avise, J. C. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* **87**, 83–93 (2006).
11. Woram, R. A. et al. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res.* **13**, 272–280 (2003).
12. Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S. & Sakaizumi, M. Evidence for different origins of sex chromosomes in closely related *Oryzias* fishes: substitution of the master sex-determining gene. *Genetics* **177**, 2075–2081 (2007).
13. White, M. J. D. *Animal Cytology and Evolution* (Univ. Press, 1973).
14. Ross, J. A. et al. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet.* **5**, e1000391 (2009).
15. Seehausen, O., van Alphen, J. J. M. & Lande, R. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**, 367–378 (1999).
16. Lande, R., Seehausen, O. & van Alphen, J. J. M. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* **112–113**, 435–443 (2001).
17. Kocher, T. D. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Rev. Genet.* **5**, 288–298 (2004).
18. Higuchi, M. & Goto, A. Genetic evidence supporting the existence of two distinct species in the genus *Gasterosteus* around Japan. *Environ. Biol. Fishes* **47**, 1–16 (1996).
19. Kitano, J., Mori, S. & Peichel, C. L. Phenotypic divergence and reproductive isolation between sympatric populations of anadromous threespine sticklebacks. *Biol. J. Linn. Soc.* **91**, 671–685 (2007).
20. Kingsley, D. M. & Peichel, C. L. In *Biology of the Three-Spined Stickleback* (eds Östlund-Nilsson, S., Mayer, I. & Huntingford, F.) 41–81 (CRC Press, 2007).
21. Peichel, C. L. et al. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr. Biol.* **14**, 1416–1424 (2004).
22. Ross, J. A. & Peichel, C. L. Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics* **179**, 2173–2182 (2008).
23. Foster, S. A. In *The Evolutionary Biology of the Threespine Stickleback* (eds Bell, M. A. & Foster, S. A.) 381–398 (Oxford Univ. Press, 1994).
24. Wootton, R. J. A *Functional Biology of Sticklebacks* (Croom Helm, 1984).
25. Kitano, J., Mori, S. & Peichel, C. L. Sexual dimorphism in the external morphology of the threespine stickleback (*Gasterosteus aculeatus*). *Copeia* **2007**, 336–349 (2007).
26. Wu, C.-I. & Davis, A. W. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *Am. Nat.* **142**, 187–212 (1993).
27. Tao, Y., Hartl, D. L. & Laurie, C. C. Sex-ratio segregation distortion associated with reproductive isolation in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 13183–13188 (2001).
28. Phadnis, N. & Orr, H. A. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* **323**, 376–379 (2009).

29. Orr, H. A. & Irving, S. Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. *Genetics* **158**, 1089–1100 (2001).
30. Sawamura, K., Roote, J., Wu, C.-I. & Yamamoto, M. T. Genetic complexity underlying hybrid male sterility in *Drosophila*. *Genetics* **166**, 789–796 (2004).
31. Gavrilov, S. *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press, 2004).
32. Arnqvist, G. & Rowe, L. *Sexual Conflict* (Princeton Univ. Press, 2005).
33. Van Ooijen, J. W. & Voorrips, R. E. *JoinMap 3.0, Software for the Calculation of Genetic Linkage Maps* (Plant Research International, 2001).
34. Van Ooijen, J. W., Boer, M. P., Jansen, R. C. & Maliepaard, C. *MapQTL 4.0, Software for the Calculation of QTL Positions on Genetic Maps* (Plant Research International, 2002).
35. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Nishitani, J. Kitajima, M. Nishida, S. Takeyama, T. Andoh, T. Kuwahara, C. Torii, Akkeshi Fisheries Cooperative Association, Akkeshi Waterfowl Center, S. Brady, A. Southwick, all members of the Peichel laboratory, and many field assistants for technical help and discussion. We thank

J. Boughman, T. Bradshaw, H. Malik, J. McKinnon, N. Phadnis and D. Schluter for comments on the manuscript. We also thank the Broad Institute for the public release of an initial stickleback genome assembly. This research was supported by the Uehara Memorial Foundation (J.K.), a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and Water and People Project of Research Institute for Humanity and Nature (S.M.), Akkeshi Town Grants-in-Aid for Scientific Research in the Lake Akkeshi-Bekanbeushi Wetland (M.K.), a Burroughs Wellcome Fund Career Award in the Biomedical Sciences (C.L.P.), and National Institutes of Health grants T32 GM07270 (J.A.R.), R01 GM071854 (C.L.P.) and P50 HG02568 (R.M.M., D.M.K. and C.L.P.).

Author Contributions J.K., S.M. and C.L.P. conceived and designed the study. F.C.J., Y.F.C., D.M.A., J.G., J.S., R.M.M. and D.M.K. contributed new reagents and carried out the SNP genotyping experiments for genome-wide linkage mapping. J.K., J.A.R., S.M., M.K. and C.L.P. performed all other experiments and analysed the data. J.K. and C.L.P. wrote the manuscript.

Author Information All SNP information has been deposited at <http://www.ncbi.nlm.nih.gov/projects/SNP/>. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.L.P. (cpeichel@fhcrc.org).

ARTICLES

Pten in stromal fibroblasts suppresses mammary epithelial tumours

Anthony J. Trimboli^{1,2*}, Carmen Z. Cantemir-Stone^{3*}, Fu Li^{1,3*}, Julie A. Wallace³, Anand Merchant³, Nicholas Creasap^{1,2}, John C. Thompson^{1,2}, Enrico Caserta^{1,2}, Hui Wang^{1,2}, Jean-Leon Chong^{1,2}, Shan Naidu^{1,2,4}, Guo Wei^{1,3}, Sudarshana M. Sharma³, Julie A. Stephens⁵, Soledad A. Fernandez⁵, Metin N. Gurcan⁶, Michael B. Weinstein^{1,2}, Sanford H. Barsky^{7†}, Lisa Yee⁸, Thomas J. Rosol⁴, Paul C. Stromberg⁴, Michael L. Robinson^{9†}, Francois Pepin^{10,11}, Michael Hallett^{10,11}, Morag Park^{10,12}, Michael C. Ostrowski^{3,13} & Gustavo Leone^{1,2,13}

The tumour stroma is believed to contribute to some of the most malignant characteristics of epithelial tumours. However, signalling between stromal and tumour cells is complex and remains poorly understood. Here we show that the genetic inactivation of *Pten* in stromal fibroblasts of mouse mammary glands accelerated the initiation, progression and malignant transformation of mammary epithelial tumours. This was associated with the massive remodelling of the extracellular matrix (ECM), innate immune cell infiltration and increased angiogenesis. Loss of *Pten* in stromal fibroblasts led to increased expression, phosphorylation (T72) and recruitment of *Ets2* to target promoters known to be involved in these processes. Remarkably, *Ets2* inactivation in *Pten* stroma-deleted tumours ameliorated disruption of the tumour microenvironment and was sufficient to decrease tumour growth and progression. Global gene expression profiling of mammary stromal cells identified a *Pten*-specific signature that was highly represented in the tumour stroma of patients with breast cancer. These findings identify the *Pten*–*Ets2* axis as a critical stroma-specific signalling pathway that suppresses mammary epithelial tumours.

Coordinated signalling between different cell types of the ‘normal stroma’ is required during embryonic and adult development¹. The stroma can be appropriately activated in response to extreme but normal physiological cues, such as wounding, inflammation or pregnancy². The stroma can also be inappropriately activated in cancer^{3,4}. In breast tumours, stromal fibroblasts are believed to adapt and continuously co-evolve with tumour epithelial cells to foster transformation and tumour growth⁵. Fibroblasts are a principal constituent of the stroma responsible for the synthesis of growth and survival factors, angiogenic and immunological chemokines, and structural components of the ECM as well as enzymes that control its turnover^{6,7}. Despite extensive evidence for a role of the tumour stroma in carcinogenesis, relatively little is known about the signalling pathways involved in the communication between the different cellular compartments of the microenvironment that contribute to the cancer phenotype.

Alterations in the phosphoinositide 3-kinase (PI3K) pathway are associated with the activation of tumour-associated stroma^{8,9}. One of the main regulators of PI3K signalling is the phosphatase and tensin homologue (PTEN), a tumour suppressor with lipid and protein phosphatase activity^{10,11}. *PTEN* inactivation disrupts multiple cellular processes associated with cell polarity, cell architecture, chromosomal integrity, cell cycle progression, cell growth and stem cell self-renewal^{12,13}. Germ-line inactivation of a single allele of *PTEN*

in both humans and mice contributes to the genesis of a variety of tumour types of epithelial origin¹⁴. Although tremendous progress in understanding PTEN function in tumour cells has been made since its discovery over a decade ago, relatively little is known about its potential role in the tumour stroma. Here, we show that *Pten* ablation in mammary stromal fibroblasts of mice results in massive remodelling of the ECM and tumour vasculature, recruitment of innate immune cells and increased malignancy of mammary epithelial tumours. Gene expression profiling of *Pten*-deleted stromal fibroblasts identified the activation of an *Ets2*-specific transcription program associated with many of these aggressive tumour phenotypes. Remarkably, the concomitant inactivation of *Ets2* in the mammary stroma reversed the increased malignancy caused by *Pten* deficiency. These findings expand *Pten*’s repertoire as a tumour suppressor by identifying the fibroblast as a key site from which it exerts its powerful tumour suppressive influence on the adjacent tumour epithelium.

Stromal *Pten* suppresses mammary epithelial tumours

To evaluate rigorously the role of *Pten* in the tumour microenvironment of breast cancer, we generated mice containing a mesenchymal-specific *Fsp-cre* transgene¹⁵ and conditional alleles of *Pten* (*Pten*^{loxP}; Supplementary Fig. 1). Cell-type-marker analysis using a β -galactosidase *Rosa26*^{LoxP} reporter allele showed specific *Fsp-cre* expression in stromal fibroblasts surrounding the mammary epithelial ducts, with

¹Department of Molecular Genetics, College of Biological Sciences, ²Department of Molecular Virology, Immunology and Medical Genetics, ³Department of Molecular and Cellular Biochemistry, College of Medicine, ⁴Department of Veterinary Biosciences, College of Veterinary Medicine, ⁵Center for Biostatistics, Office of Health Sciences, ⁶Department of Biomedical Informatics, ⁷Department of Pathology and ⁸Department of Surgery, School of Medicine, The Ohio State University, Columbus, Ohio 43210, USA. ⁹Center for Molecular and Human Genetics, Columbus Children’s Research Institute, Columbus, Ohio 43205, USA. ¹⁰Department of Biochemistry, Rosalind and Morris Goodman Cancer Center, ¹¹McGill Center for Bioinformatics, ¹²Department of Oncology, McGill University, Québec H3A 1A1, Canada. ¹³Tumor Microenvironment Program, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio 43210, USA. [†]Present addresses: Department of Pathology, University of Nevada School of Medicine, Reno, Nevada 89557, and Nevada Cancer Institute, Las Vegas, Nevada 89135, USA (S.H.B.); Department of Zoology, Miami University, Oxford, Ohio 45056, USA (M.L.R.).

*These authors contributed equally to this work.

no expression in cytokeratin-positive epithelial cells, F4/80-positive macrophages and CD31-positive endothelial cells (Fig. 1a and Supplementary Fig. 2a, b). Western blot and PCR assays demonstrated efficient cre-mediated deletion of *Pten*^{loxP} in stromal fibroblasts isolated from *Fsp-cre;Pten*^{loxP/loxP} mammary glands (Fig. 1b and Supplementary Fig. 3a). Examination of mammary sections by immunohistochemistry (IHC) and immunofluorescence showed deletion of *Pten*^{loxP} that was confined to stromal fibroblasts, with no collateral deletion in epithelial ducts or the adjacent myoepithelium (Fig. 1c and Supplementary Fig. 3b, c). Interestingly, this resulted in the expansion of the ECM, but did not lead to the transformation of the mammary epithelium (Fig. 1c, e).

We then examined the role of stromal *Pten* on mammary tumorigenesis using an established mouse model of breast cancer, *MMTV-ErbB2/neu* (*ErbB2*)¹⁶. To avoid possible confounding effects caused by *Pten* deletion in mesenchymal cells of other organs, mammary glands from *Fsp-cre;Pten*^{loxP/loxP}, *ErbB2;Pten*^{loxP/loxP} and *ErbB2;Fsp-cre;Pten*^{loxP/loxP} donors were transplanted into syngeneic wild-type recipients¹⁷ and tumour development was monitored over the course of several months. By genetically marking the stroma with the *Rosa26*^{LoxP} reporter allele, we demonstrated that both the epithelium and its associated stroma were effectively transplanted into host female mice (Supplementary Fig. 4). Loss of *Pten* in stromal fibroblasts dramatically increased the incidence of *ErbB2*-driven mammary tumours (Fig. 1d–f). By 16 weeks post-transplantation, these lesions progressed to adenoma, carcinoma *in situ* and invasive carcinoma (Fig. 1g), and by 26 weeks most females met the criteria for early removal due to excessive tumour burden (Fig. 1f). Histological examination showed that *ErbB2*-tumour cells in *Pten* stromal-deleted tumours retained their typical oncogene-specific morphology, with

small nuclei, fine chromatin and abundant eosinophilic cytoplasm¹⁸. In contrast to non-deleted tumours^{18,19}, *Pten* stromal-deleted tumours had a significant amount of stroma surrounding and infiltrating the epithelial masses (Fig. 1g). PCR-based and immunohistochemical assays confirmed that tumours had intact *Pten*^{loxP} alleles in the epithelial compartment (Supplementary Fig. 5a, b and data not shown). Moreover, we used the *Rosa26*^{LoxP} reporter allele to mark genetically early epithelial to mesenchymal transition events¹⁵ and found no evidence of epithelial to mesenchymal transition in tumours that either contained or lacked *Pten* in stromal fibroblasts (data not shown). Thus, the analysis of the *ErbB2* breast-cancer tumour model identified a potent tumour suppressor role for *Pten* in stromal fibroblasts of the mammary gland.

Stromal *Pten* controls ECM and innate immune functions

To investigate the tumour suppressive mechanism of *Pten* action in stromal fibroblasts, we profiled the transcriptome of mammary stromal fibroblasts isolated from *Pten*^{loxP/loxP} and *Fsp-cre;Pten*^{loxP/loxP} females. Details of sample collection, processing of Affymetrix oligo-arrays and expression data are available in Methods. Briefly, we implemented class comparison analyses of all probe sets on the Affymetrix mouse genome 430 2.0 array to identify genes differentially expressed between the two genetic groups. We also used an unbiased approach similar to gene set enrichment analysis²⁰ to identify *a priori* defined groups of genes that were significantly differentially expressed. The analysis of over 14,000 mouse genes identified 129 upregulated and 21 downregulated unique genes in response to *Pten* deletion (Supplementary Fig. 6a, b; greater than fourfold at $P < 0.001$; Supplementary Tables 1 and 2). Reverse transcription followed by quantitative PCR (quantitative RT-PCR) assays on a subset of genes confirmed more than 85% of these expression

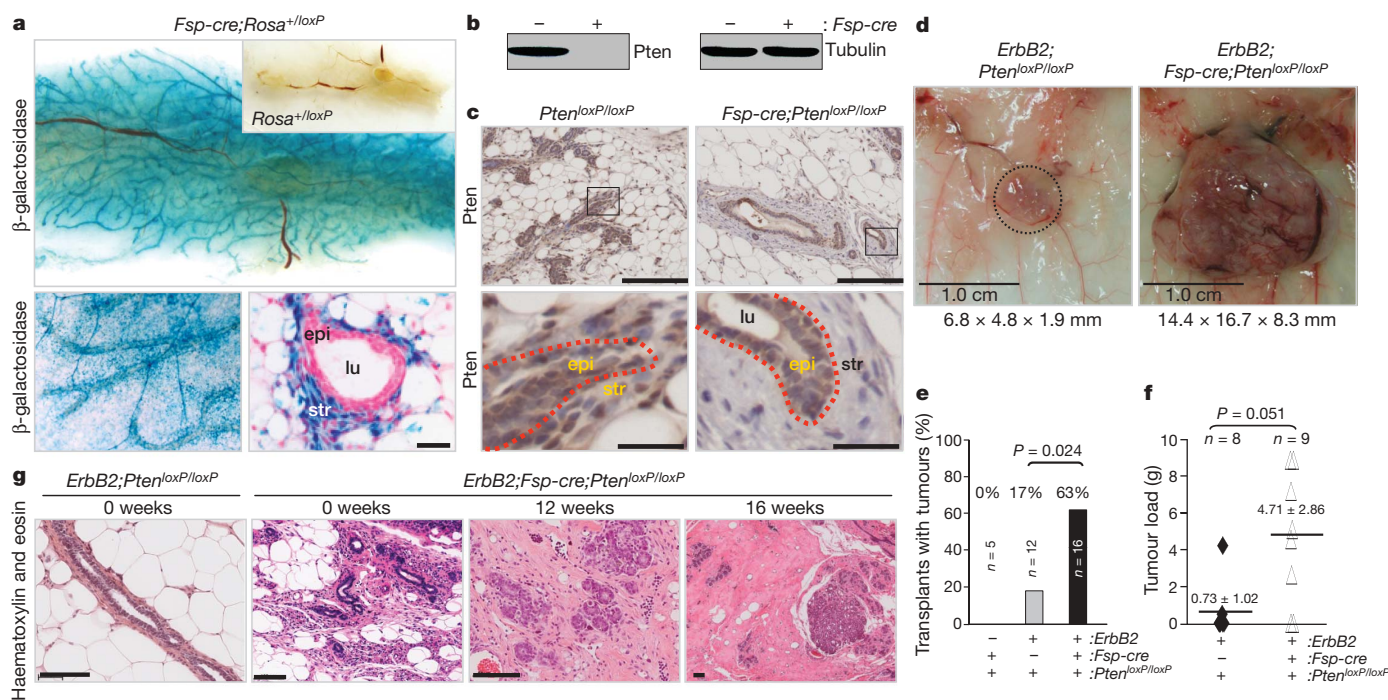


Figure 1 | Stromal fibroblast-specific deletion of *Pten*. **a**, Wholemount, X-gal-stained mammary glands from *Fsp-cre;Rosa*^{+/loxP} and *Rosa*^{+/loxP} (top, inset) mice. Higher magnification of wholemount gland (bottom left) and a histological cross section (bottom right); scale bar, 30 μ m. lu, Lumen; epi, epithelium; str, stroma. **b**, Representative western blot analysis of mammary fibroblast lysates derived from 8-week-old *Pten*^{loxP/loxP} mice with (+) or without (–) *Fsp-cre*. **c**, Paraffin sections from 8-week-old female mammary glands stained with a *Pten*-specific antibody; lower panels represent higher magnifications of boxed areas; scale bars: top panels, 200 μ m; bottom panels, 30 μ m. lu, Lumen; epi, epithelial compartment; str, stromal compartment; red dotted line indicates the border between the two compartments.

d, Tumours collected at 26 weeks post-transplantation. **e**, Tumour development by 16 weeks in mammary glands with the indicated genotypes. Tumorigenicity was determined by palpation or histological presentation of adenoma/carcinoma at each implantation site and statistically analysed using Fisher's exact test. n, Total number of transplants. **f**, Total tumour burden at 26 weeks post-transplantation in mammary glands with the indicated genotypes. Values represent mean \pm s.d. Differences were tested using the non-parametric Wilcoxon rank sum test. **g**, H&E-stained sections of mammary glands harvested at time of transplantation (0 weeks) and indicated times post-transplantation; scale bars, 100 μ m.

changes using independent fibroblast samples (Supplementary Fig. 6c and Supplementary Table 3). Fibroblast samples used to probe the oligo-arrays lacked expression of macrophage-, endothelial- and epithelial-specific genes, confirming the purity of these fibroblast preparations (Supplementary Fig. 6d). Functional annotation^{21,22} of *Pten*-responsive targets revealed a remarkable bias towards genes encoding proteins involved in ECM remodelling, wound healing and chronic inflammation^{21,22} (Fig. 2a and Supplementary Tables 1 and 2). Given this unexpected convergence of function, we performed a more thorough cellular and molecular analysis of *Pten*-deleted stroma. Staining of consecutive mammary gland sections with haematoxylin and eosin (H&E) and Mason's trichrome stains indicated enhanced deposition of collagen in *Pten*-deleted stroma, which was independent of *ErbB2*-oncogene expression (Fig. 2b, c and Supplementary Fig. 7a). IHC and western blot assays using collagen type-specific antibodies showed that the non-cellular material consisted mostly of type-I collagen and not the basement membrane type-IV collagen (Fig. 2b, c and Supplementary Fig. 7b, c). There was significant infiltration of F4/80-positive macrophages into stromal *Pten*-deleted mammary glands (Fig. 2d, e), and this was independent of *ErbB2* expression (Supplementary Fig. 8a). The abundance of B and T cells did not change in response to stromal deletion of *Pten* (data not shown). From these experiments, we conclude that ablation of *Pten* in stromal fibroblasts recapitulates two key events associated with tumour malignancy: increased ECM deposition and innate immune cell infiltration.

Stromal *Pten* loss activates *Ets2*

Along with the remarkable remodelling of the tumour microenvironment, loss of *Pten* in stromal fibroblasts resulted in activation of the Ras, JNK and Akt pathways. Western blot analysis using protein lysates derived from *Pten*-deleted stromal fibroblasts demonstrated an increase in the phospho-specific forms of Akt (Akt^{T308/S473}) and JNK (JNK^{T183/Y185}) (Fig. 2f and Supplementary Fig. 8b). Immunohistochemical assays confirmed the activation of Akt and JNK in stromal fibroblasts, and, interestingly, revealed a profound activation of these two pathways in ductal epithelial cells adjacent to the *Pten*-deleted stroma (Fig. 2g and Supplementary Fig. 8b). This analysis also showed increased levels of phospho-Erk1/2 in *Pten*-deleted stromal fibroblasts; however, this increase could not be detected in primary cultured fibroblasts (Fig. 2f, g), presumably owing to the constitutive *Pten*-independent activation of Erk1/2 by serum-stimulation²³.

Among the many expression changes observed in *Pten*-deleted stromal fibroblasts we noted that there was a significant increase of *Ets2* messenger RNA (mRNA) levels (2.8-fold, $P < 0.001$). This induction is notable because the *Ets2* transcription factor is known to be transcriptionally induced by MAPK^{24,25} activation and its function to be post-translationally enhanced by the Akt- and JNK-mediated phosphorylation of its pointed domain at threonine 72 (*Ets2*^{T72})^{23,26}. We confirmed the higher levels of *Ets2* mRNA and protein in *Pten*-deleted fibroblasts (approximately threefold, $P < 0.001$; Supplementary Fig. 9a, b) and, consistent with the activation of Akt and JNK in these mammary glands, there was a marked increase of phospho-*Ets2*^{T72} in stromal fibroblasts and adjacent epithelial ducts (Fig. 2h, i). Loss of *Pten* in stromal fibroblasts resulted in the induction of several genes involved in ECM remodelling and macrophage recruitment, two of which, *Mmp9* and *Ccl3*, are known to be direct transcriptional targets of *Ets2*^{27,28} (Supplementary Figs 6c and 9c). The increase of *Mmp9* expression appears to be of pathological relevance because *in situ* zymography²⁹ showed robust *Mmp9* activity in tumour samples (Supplementary Fig. 9d). Chromatin immunoprecipitation (ChIP) assays showed an increase in the loading of *Ets2* onto the *Mmp9* and *Ccl3* promoters in *Pten*-deleted mammary fibroblasts (Supplementary Fig. 9e), suggesting a direct role for *Ets2* in the transcriptional regulation of these two target genes *in vivo*. Together, these data illustrate the extensive molecular reprogramming that takes place in the tumour and its microenvironment in response to ablation of *Pten* in stromal fibroblasts.

Stromal *Ets2* promotes mammary tumorigenesis

To determine whether *Ets2* promotes a microenvironment conducive to tumour growth, we analysed the consequences of ablating a conditional allele of *Ets2* (*Ets2*^{loxP})³⁰ in mammary stromal fibroblasts of a well-characterized mouse model of breast cancer, *MMTV-PyMT* (*PyMT*)³¹. The *PyMT* oncogene initiates the rapid onset and progression of mammary tumours and thus represents an appropriate model for evaluating any potential delay that loss of *Ets2* might have on tumorigenesis. The efficient *Fsp-cre*-mediated ablation of *Ets2* in stromal fibroblasts was facilitated by using mice carrying conventional and conditional knockout alleles of *Ets2* (DNA binding domain *Ets2*^{db/loxP})³² (Fig. 3a and Supplementary Fig. 10a, b). Ablation of *Ets2* in these cells had no detectable physiological consequence on the development of mammary glands, either during puberty or pregnancy (M.C.O., unpublished observations). The evaluation of *PyMT*;*Fsp-cre*;*Ets2*^{db/loxP} and control *PyMT*;*Ets2*^{db/loxP} mice over a period of three months showed that ablation of *Ets2* in mammary fibroblasts significantly reduced the tumour load (Fig. 3b) and slowed progression to adenoma and early carcinoma (Fig. 3c). Quantitative RT-PCR showed high levels of *Mmp9* expression in tumour-associated fibroblasts containing *Ets2* and low levels in *Ets2*-deleted fibroblasts (Fig. 3e). Because *Mmp9* activity is known to mediate the release of matrix-bound VEGF-A to its active isoforms, including VEGF₁₆₄ (ref. 33), we visualized the spatial distribution of VEGF₁₆₄ and *Mmp9* activity by immunofluorescence. These assays showed that the accumulation of VEGF₁₆₄, which was particularly acute within collagen-1A-rich stromal locations overlapping *Mmp9* activity, was significantly decreased in stromal-deleted *Ets2* tumours (Fig. 3d, f). Given that VEGF₁₆₄ is a specific ligand for VEGF Receptor 2 (VEGFR2; FLK-1; KDR), one of the most potent mediators of VEGF-induced endothelial signalling and angiogenesis³⁴, we also evaluated VEGFR2 status by immunostaining tumour sections with antibodies specific for CD31 and the phospho-activated form of the murine VEGF receptor (VEGFR2^{Y1173})³⁵. This analysis revealed a fourfold decrease in the number of CD31/VEGFR2^{Y1173} double-positive cells in *Ets2*-deleted versus non-deleted tumour samples (Fig. 3f, g). Together, these data show that loss of *Ets2* in stromal fibroblasts results in decreased *Mmp9* activity in the tumour ECM and reduced VEGFR2^{Y1173} activation in the tumour vasculature.

Ets2 is a key component of the stromal *Pten* axis

We then entertained the hypothesis that *Ets2* may be contributing to the remodelling of the tumour microenvironment caused by stromal deletion of *Pten*. To test this possibility directly, we compared tumour incidence in *Pten*^{loxP/loxP}, *Fsp-cre*;*Pten*^{loxP/loxP} and *Fsp-cre*;*Pten*^{loxP/loxP};*Ets2*^{db/loxP} mammary glands that were orthotopically injected with an established *ErbB2*-initiated mammary tumour cell line (NT2.5)³⁶. This orthotopic model recapitulated the consequences of deleting *Pten* in the mammary stroma that were observed in the genetically engineered *ErbB2*-mouse model described earlier in this study. Indeed, tumour incidence and tumour load in injected *Fsp-cre*;*Pten*^{loxP/loxP} females was markedly higher than in control *Pten*^{loxP/loxP} females (Fig. 4a, b). Importantly, mammary glands doubly deleted for stromal *Pten* and *Ets2* had fewer and smaller tumours than glands deleted for *Pten* only. These mammary tumours had decreased numbers of macrophages and recruitment of new vasculature (Fig. 4c–f). Loss of *Pten* and *Ets2*, however, failed to reduce the tumour load and collagen deposition fully to control levels (Fig. 4b and Supplementary Fig. 11a, b), suggesting that additional effectors must contribute towards *Pten*'s tumour suppressor functions. From these data, we conclude that *Ets2* is a major component of the *Pten* tumour-suppressive axis that acts in the stromal fibroblast compartment of mammary glands.

Human breast tumour stroma contains a *Pten* signature

To determine the relevance of these findings to human breast cancer, we compared the mouse stromal fibroblast *Pten* expression signature

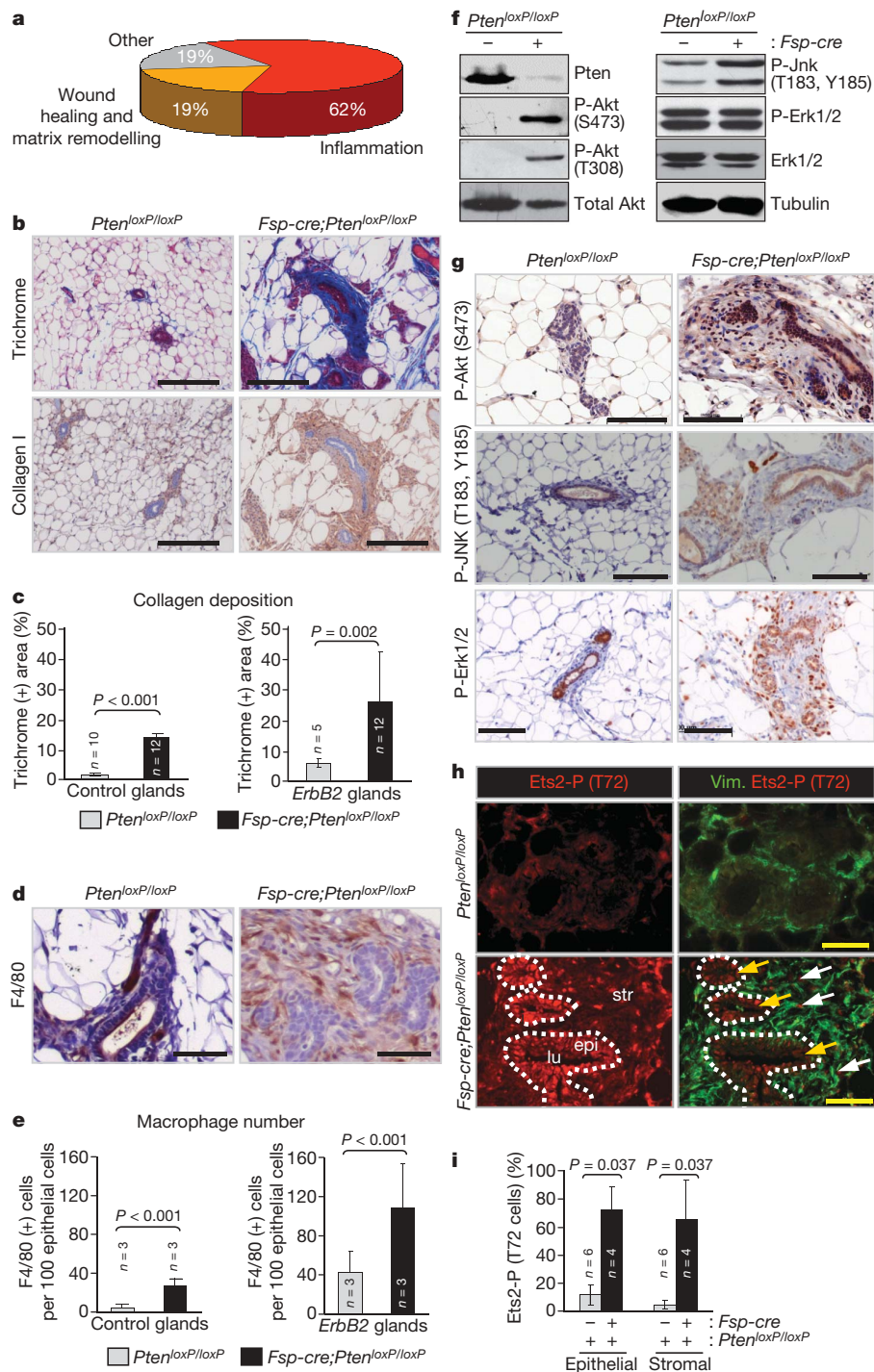


Figure 2 | Characterization of ECM deposition and immune cell infiltration.

a, Schematic representation of the biological processes affected by differentially expressed genes (more than fourfold) in *Pten*-deleted stromal fibroblasts. **b**, Mammary gland paraffin sections stained with Masson's trichrome and collagen-I-specific antibodies, respectively; scale bars, 200 μ m. **c**, Trichrome-stained sections were quantified for collagen deposition; mammary glands in the absence (left) or presence (right) of *ErbB2* were analysed, respectively. Values are means \pm s.d.; the Wilcoxon rank sum test was used to compare groups. **d**, Mammary gland paraffin sections stained with the macrophage-specific marker F4/80; scale bars, 50 μ m. **e**, Quantification of F4/80-positive stained stromal cells in mammary glands in the absence (left) or presence (right) of *ErbB2*, respectively. Values are means \pm s.d.; the Wilcoxon rank sum test was used to compare groups.

f, Western blot analysis of whole-cell lysates derived from mammary stromal fibroblasts. **g**, Mammary gland paraffin sections stained with phospho-Akt^{S473}, phospho-JNK^{T183/Y185} and phospho-Erk1/2-specific antibodies; scale bars, 100 μ m. All analyses were performed using tissue or cells from 8-week-old females. **h**, Frozen mammary tissue sections stained with vimentin (green) and phospho-Ets2^{T72}-specific (red) antibodies. Note that loss of *Pten* in the mammary stroma increased Ets2 phosphorylation in both the stromal and epithelial compartments. Dotted white line indicates the stromal–epithelial boundary. lu, Lumen; epi, epithelium; str, stroma; scale bars, 50 μ m. **i**, Quantification of mammary epithelial and stromal cells that stained positive for nuclear phospho-Ets2^{T72}. Values are means \pm s.d.; the Wilcoxon rank sum test was used to compare groups.

to the expression signatures derived from laser-captured tumour stroma (49 samples) and adjacent normal stroma (52 samples) in patients with breast cancer³⁷. This analysis identified 137 human

orthologues from the 150 differentially expressed mouse genes detected by the Affymetrix oligo-arrays shown in Supplementary Fig. 6a, b. Of these 137 orthologues, 129 genes were represented in

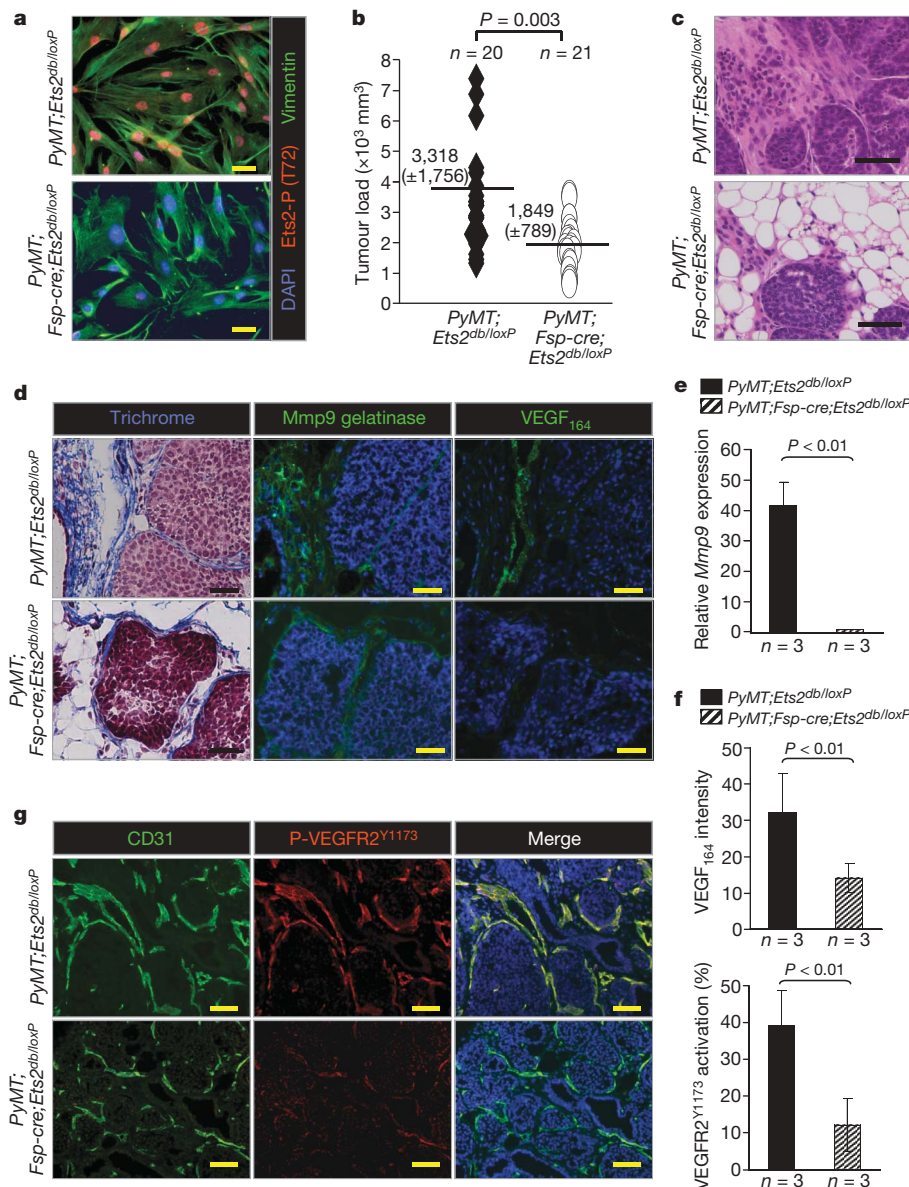


Figure 3 | *Ets2* ablation in stromal fibroblasts restricts mammary tumorigenesis. **a**, Immunofluorescence staining of cultured mammary fibroblasts with vimentin (green), phospho-Ets2^{T72} (red) antibodies and counterstained with 4,6-diamidino-2-phenylindole (DAPI) (blue). **b**, Total mammary tumour volume of *PyMT*; *Ets2*^{db/loxP} (*n* = 20) and *PyMT*; *Fsp-cre*; *Ets2*^{db/loxP} (*n* = 21) mice collected 30 days after tumour initiation. Values represent the mean \pm s.d. **c**, H&E staining of tumours harvested from *PyMT*; *Ets2*^{db/loxP} or *PyMT*; *Fsp-cre*; *Ets2*^{db/loxP} mice. **d**, Consecutive sections stained for (left to right): trichrome, Mmp9 gelatinase activity and VEGF₁₆₄, and counterstained with DAPI from *PyMT*; *Ets2*^{db/loxP} and *PyMT*; *Fsp-cre*; *Ets2*^{db/loxP} mammary tumours. **e**, Quantification of *Mmp9* mRNA

expression by quantitative RT-PCR. **f**, Quantification of VEGF₁₆₄ immunofluorescence staining in tumour stroma (top) and tumour endothelial cells co-expressing CD31 and phospho-VEGFR2^{Y1173} (bottom). **g**, Tumour vascular endothelial cells visualized by immunofluorescence double staining with CD31 (green) and p-VEGFR2^{Y1173} (red), and counterstained with DAPI (blue) in mammary tumours collected 1 week after tumour initiation in *PyMT*; *Ets2*^{db/loxP} and *PyMT*; *Fsp-cre*; *Ets2*^{db/loxP} mice. All analyses used tissue or cells from 9- to 10-week-old females; scale bars, 50 μ m. In **e** and **f**, bars represent means and error bars are s.d. Student's *t*-test was used for all the statistical comparisons between groups.

the expression platform used (Agilent) for the analysis of human patient stroma samples³⁷. Only 70 of these 129 genes had highly variable gene expression across all human stromal samples (a variance cutoff of greater than 0.5). The heat map generated for the human stroma data set showed that this 70 gene subset derived from the mouse *Pten*-signature was sufficient to distinguish normal from tumour stroma in all patients (Fig. 5a; $P = 8 \times 10^{-5}$ as determined by a permutation test). Principal component analysis also discriminated normal from tumour stroma perfectly ($P < 1 \times 10^{-10}$; Supplementary Fig. 12). Interestingly, 12 of the 70 human orthologues identified by the *Pten*-signature (Fig. 5a: gene names highlighted in red; Supplementary Fig. 13) were previously shown to be differentially expressed in the tumour stroma of patients with breast cancer and to

be associated with recurrence³⁷ (Fig. 5b; $P = 2.5 \times 10^{-8}$, Fisher's exact analysis). These analyses suggest that the fibroblast *Pten*-expression signature identified by our stromal mouse model represents a significant subset of the total gene signature expressed in the stroma of human breast cancer. We interpret these results to mean that a portion of the transcriptome regulated by *Pten* in mammary stromal fibroblasts is dysregulated in the tumour stroma of patients with breast cancer.

We also evaluated the relevance of the *Pten*–*Ets2* relationship in human cancer by immunohistochemical staining of breast-cancer tissue microarrays (TMAs) with antibodies specific for *Pten*, *P-Ets2*^{T72} and *P-Akt*^{S473} (Fig. 5c). From the analysis of 99 patient samples with invasive carcinoma, PTEN expression was scored as

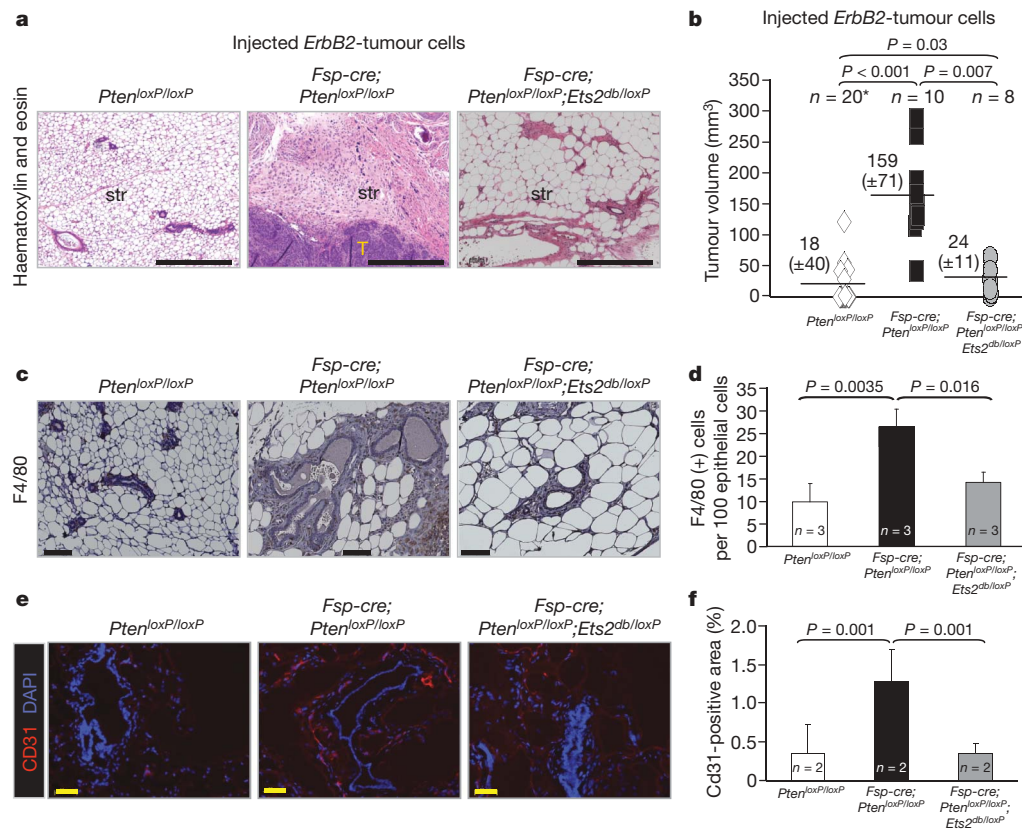


Figure 4 | Loss of *Ets2* in stromal fibroblasts diminishes tumour growth in stromal *Pten*-deleted mammary glands. **a**, H&E sections of mammary glands after orthotopic injection of the *ErbB2*-expressing tumour cell line NT2.5. T, tumour; str, stroma; scale bars, 500 μ m. **b**, Volumes of tumours collected 21 days after injection. *Pten*^{loxP/loxP}; *Ets2*^{db/loxP} ($n = 10$) and *Pten*^{loxP/loxP} ($n = 10$) control groups were combined (*, $n = 20$) after it was determined that there was no statistical difference in tumour incidence or load between these two control groups. Values are means \pm s.d. **c**, Sections from mammary glands with the indicated genotypes stained with the

absent or low in approximately 50% of samples. Importantly, PTEN staining in the TMAs was negatively correlated with P-AKT^{S473} and nuclear P-ETS2^{T72}, whereas P-AKT^{S473} and nuclear P-ETS2^{T72} showed a positive association (Fig. 5d). These results suggest that activation of P-ETS2^{T72} in human breast-cancer stroma is a pathological event that is favoured by a reduction in PTEN expression.

Discussion

Histopathology and molecular studies suggest that malignant tumours consist of a complex cellular system that is dependent on reciprocal signalling between tumour cells and the adjacent stroma. However, the signalling pathways that mediate the communication between the various cell types in the tumour remain virtually unknown. We recently developed a mesenchymal-specific *cre* mouse¹⁵ and used it here to examine the consequences of inactivating *Pten* in mammary stromal fibroblasts. Using this system we show, for the first time to our knowledge, that *Pten* in stromal fibroblasts has a critical role in the suppression of epithelial mammary tumours that is, in part, mediated through an *Ets2*-regulated transcriptional program.

The tumour suppressor functions of PTEN have been extensively studied in the tumour cell^{38–40}. We show here that genetic ablation of *Pten* in mammary stromal fibroblasts of mice alters the expression profile of these cells to increase ECM, chemokine and cytokine production in the tumour microenvironment. As a result, *Pten* stromal-deleted tumours exhibit high levels of collagen, macrophage recruitment and vascular networks, which together favour the initiation and progression of mammary epithelial tumours. Remarkably, side-by-side evaluation of histopathology by independent pathologists could not distinguish

macrophage-specific marker F4/80. Scale bars, 100 μ m. **d**, Quantification of stromal cells positive for F4/80 in mammary glands. **e**, Frozen mammary gland sections stained with the endothelial-specific antibody, CD31. Scale bars, 50 μ m. **f**, Quantification of CD31-positive staining. In **d** and **f**, bars represent means and error bars are s.d. For all the statistical analyses, an analysis of variance model with Bonferroni adjustment was used. Pairwise comparisons shown have a significant difference between marked genetic groups.

tumours between *Pten* stromal-deleted mice and human patients with breast cancer, highlighting the importance of modelling stromal cell compartments of the tumour microenvironment. The mechanism by which *Pten* in the stroma likely exerts its tumour suppressor role is through the control of multiple signalling pathways, including components of the Ras, Akt and JNK networks, which together culminate in the regulation of *Ets2* transcriptional activity. The fact that loss of *Ets2* in mammary stromal fibroblasts diminished the oncogenic consequences of deleting *Pten* in these cells underscores the importance of the stromal *Pten*–*Ets2* axis in stromal fibroblasts during tumour suppression. These observations are consistent with previous work from Oshima and colleagues that showed a critical cell non-autonomous role for *Ets2* in the growth of mammary tumours in mice⁴¹ and with the identification of *Ets2* activation as a key event associated with breast cancer in human patients having poor prognosis^{42–44}. The relevance of this mouse *Pten*–*Ets2* tumour suppressor axis to breast cancer is underlined by the high correspondence between the mouse and human stromal expression signatures. The observation that the dire consequences of targeting this *Ets2*-driven stromal program are tumour-specific, sparing normal mammary development, emphasizes the potential use of stromal-specific strategies for therapeutic intervention in human breast cancer.

In summary, this work identifies *Pten*–*Ets2* as a key regulatory axis in stromal fibroblasts that suppresses mammary epithelial tumours by profoundly attenuating some of the most malignant characteristics of the tumour microenvironment. This novel function of *Pten* may be relevant in the suppression of epithelial tumours of other organs, but may also extend beyond cancer, to conditions where the microenvironment may impact disease manifestation, such as in autoimmune

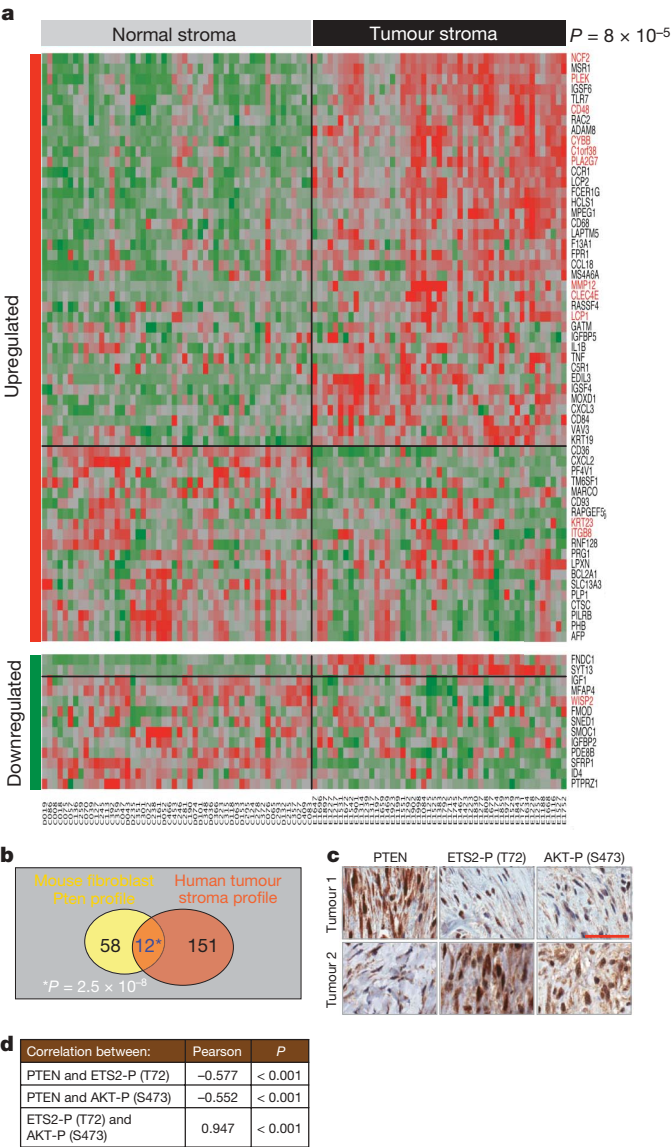


Figure 5 | Pten-signature is represented in breast cancer stroma. **a**, Heat map displaying the expression of the human orthologues of the 70-gene subset in normal- and tumour-stroma from human patients with breast cancer. The 70-gene subset derived from the mouse Pten-signature includes 57 genes upregulated (red bar on y-axis) and 13 genes downregulated (green bar on y-axis). Red and green regions inside the heat map indicate relative gene expression levels (red, increased; green, decreased). The *P* value indicates the ability of the mouse 70-gene signature to partition normal and tumour stroma in patients with breast cancer (see Statistical Methods). The coded patient identities are listed at the bottom. **b**, Venn diagram depicting the overlap between the mouse Pten-deleted fibroblast and human stroma microarray data sets. The 12 genes highlighted in red (**a**, right margin) are common between the mouse Pten 70-gene signature and the human 163-gene signature that has been shown to associate with recurrence³⁸. This overlap is highly significant ($P = 2.5 \times 10^{-8}$; Fisher's exact test). **c**, Representative PTEN, P-ETS2^{T72} and P-AKT^{S473} IHC staining in human breast carcinoma samples from the tissue microarray. Scale bar, 50 μ m. **d**, Pearson correlations between PTEN, P-ETS2^{T72} and P-AKT^{S473} expression based on Allred scores of a tissue microarray (**c**) containing 99 patients with advanced breast carcinoma.

syndromes⁴⁵, lung fibrosis⁴⁶ and neurodegeneration⁴⁷. Interestingly, the stromal Pten expression signature identified here includes genes that have been causally linked to ECM deposition and inflammation in rheumatoid arthritis, lung fibrosis and neurodegeneration (Supplementary Tables 1 and 2). These data offer a molecular basis for how altered Pten signalling in the tumour stroma may elicit broad

responses in a variety of cells in the tumour microenvironment that contribute to disease manifestation.

METHODS SUMMARY

Transgenic mice. Generation of *Fsp-cre* mice has been described¹⁵. *Pten*^{loxP} mice were created following the strategy described in Supplementary Fig. 1. *Ets2*^{loxP} mice were generated by standard techniques³⁰. Animals were maintained and killed following institutional guidelines. Tenth-generation congenic (N10) FVB/N animals were used for transplantation and orthotopic injection studies.

Tissue processing, histology. Tissues were either embedded in OCT or fixed (4% paraformaldehyde or formalin) and embedded in paraffin. Frozen sections were used for X-gal staining as previously described¹⁵. IHC on tissue microarray sections from 99 patients with breast carcinoma was scored using Allred score's system⁴⁸.

Isolation of primary mammary fibroblasts. Primary mammary fibroblasts were purified following the protocol published previously with minor modifications⁴⁹. Mammary glands were dissected from 8-week-old female mice, minced and digested with collagenase (0.15% Collagenase I, 160 U ml⁻¹ hyaluronidase, 1 μ g ml⁻¹ hydrocortisone and 10 μ g ml⁻¹ insulin with penicillin and streptomycin) in a 5% CO₂ incubator overnight at 37 °C. Collagenase was neutralized with 10% FBS-DMEM medium. Digested tissue was resuspended in medium and subjected to gravity for 12–15 min. Pellets were washed three times to collect epithelial organoids, and supernatants were subjected four more times to gravity sedimentation and then cultured.

RNA and microarray analysis. RNA was harvested with Trizol according to the manufacturer's instructions (Invitrogen). RNA quality and concentration were assessed with Bioanalyser and Nanodrop RNA 6000 nano-assays. RNA samples were hybridized to Affymetrix GeneChip Mouse genome 430 2.0 platform at the Microarray Shared Resource Facility, Ohio State University Comprehensive Cancer Center. The microarray data were deposited with Gene Expression Omnibus (GEO) and can be viewed at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=zhmbrcoqaacyite&acc=GSE16073>.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 May; accepted 7 September 2009.

1. Wiseman, B. S. & Werb, Z. Stromal effects on mammary gland development and breast cancer. *Science* **296**, 1046–1049 (2002).
2. Nelson, C. M. & Bissell, M. J. Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu. Rev. Cell Dev. Biol.* **22**, 287–309 (2006).
3. Mueller, M. M. & Fusenig, N. E. Friends or foes – bipolar effects of the tumour stroma in cancer. *Nature Rev. Cancer* **4**, 839–849 (2004).
4. Schedin, P. Pregnancy-associated breast cancer and metastasis. *Nature Rev. Cancer* **6**, 281–291 (2006).
5. Littlepage, L. E., Egeblad, M. & Werb, Z. Coevolution of cancer and stromal cellular responses. *Cancer Cell* **7**, 499–500 (2005).
6. Bhowmick, N. A., Neilson, E. G. & Moses, H. L. Stromal fibroblasts in cancer initiation and progression. *Nature* **432**, 332–337 (2004).
7. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nature Rev. Cancer* **6**, 392–401 (2006).
8. Cully, M., You, H., Levine, A. J. & Mak, T. W. Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nature Rev. Cancer* **6**, 184–192 (2006).
9. Bergamaschi, A. et al. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *J. Pathol.* **214**, 357–367 (2008).
10. Myers, M. P. et al. The lipid phosphatase activity of PTEN is critical for its tumor suppressor function. *Proc. Natl Acad. Sci. USA* **95**, 13513–13518 (1998).
11. Stambolic, V. et al. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* **95**, 29–39 (1998).
12. Salmena, L., Carracedo, A. & Pandolfi, P. P. Tenets of PTEN tumor suppression. *Cell* **133**, 403–414 (2008).
13. Knobbe, C. B., Lapin, V., Suzuki, A. & Mak, T. W. The roles of PTEN in development, physiology and tumorigenesis in mouse models: a tissue-by-tissue survey. *Oncogene* **27**, 5398–5415 (2008).
14. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. *Pten* is essential for embryonic development and tumour suppression. *Nature Genet.* **19**, 348–355 (1998).
15. Trimboli, A. J. et al. Direct evidence for epithelial–mesenchymal transitions in breast cancer. *Cancer Res.* **68**, 937–945 (2008).
16. Guy, C. T. et al. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc. Natl Acad. Sci. USA* **89**, 10578–10582 (1992).
17. Cases, S. et al. Development of the mammary gland requires DGAT1 expression in stromal and epithelial tissues. *Development* **131**, 3047–3055 (2004).

18. Andrechek, E. R. *et al.* Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. *Proc. Natl Acad. Sci. USA* **97**, 3444–3449 (2000).
19. Desai, K. V. *et al.* Initiating oncogenic event determines gene-expression patterns of human breast cancer models. *Proc. Natl Acad. Sci. USA* **99**, 6967–6972 (2002).
20. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
21. Dennis, G., Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).
22. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
23. Weng, L. P., Brown, J. L., Baker, K. M., Ostrowski, M. C. & Eng, C. PTEN blocks insulin-mediated ETS-2 phosphorylation through MAP kinase, independently of the phosphoinositide 3-kinase pathway. *Hum. Mol. Genet.* **11**, 1687–1696 (2002).
24. Fowles, L. F. *et al.* Persistent activation of mitogen-activated protein kinases p42 and p44 and ets-2 phosphorylation in response to colony-stimulating factor 1/c-fms signaling. *Mol. Cell. Biol.* **18**, 5148–5156 (1998).
25. McCarthy, S. A. *et al.* Rapid phosphorylation of Ets-2 accompanies mitogen-activated protein kinase activation and the induction of heparin-binding epidermal growth factor gene expression by oncogenic Raf-1. *Mol. Cell. Biol.* **17**, 2401–2412 (1997).
26. Smith, J. L. *et al.* ets-2 is a target for an akt (Protein kinase B)/jun N-terminal kinase signaling pathway in macrophages of motheaten-viable mutant mice. *Mol. Cell. Biol.* **20**, 8026–8034 (2000).
27. Watabe, T. *et al.* The Ets-1 and Ets-2 transcription factors activate the promoters for invasion-associated urokinase and collagenase genes in response to epidermal growth factor. *Int. J. Cancer* **77**, 128–137 (1998).
28. Wei, G. *et al.* Activated Ets2 is required for persistent inflammatory responses in the motheaten viable model. *J. Immunol.* **173**, 1374–1379 (2004).
29. Ludwig, T. Local proteolytic activity in tumor cell invasion and metastasis. *Bioessays* **27**, 1181–1191 (2005).
30. Wei, G. *et al.* Ets1 and Ets2 are required for endothelial cell survival during embryonic angiogenesis. *Blood* **114**, 1123–1130 (2009).
31. Lin, E. Y. *et al.* Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am. J. Pathol.* **163**, 2113–2126 (2003).
32. Yamamoto, H. *et al.* Defective trophoblast function in mice with a targeted mutation of Ets2. *Genes Dev.* **12**, 1315–1326 (1998).
33. Lee, S., Jilani, S. M., Nikolova, G. V., Carpizo, D. & Iruela-Arispe, M. L. Processing of VEGF-A by matrix metalloproteinases regulates bioavailability and vascular patterning in tumors. *J. Cell Biol.* **169**, 681–691 (2005).
34. Millauer, B. *et al.* High affinity VEGF binding and developmental expression suggest Flk-1 as a major regulator of vasculogenesis and angiogenesis. *Cell* **72**, 835–846 (1993).
35. Sakurai, Y., Ohgimoto, K., Kataoka, Y., Yoshida, N. & Shibuya, M. Essential role of Flk-1 (VEGF receptor 2) tyrosine residue 1173 in vasculogenesis in mice. *Proc. Natl Acad. Sci. USA* **102**, 1076–1081 (2005).
36. Dakappagari, N. K. *et al.* Conformational HER-2/neu B-cell epitope peptide vaccine designed to incorporate two native disulfide bonds enhances tumor cell binding and antitumor activities. *J. Biol. Chem.* **280**, 54–63 (2005).
37. Finak, G. *et al.* Stromal gene expression predicts clinical outcome in breast cancer. *Nature Med.* **14**, 518–527 (2008).
38. Berns, K. *et al.* A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* **12**, 395–402 (2007).
39. Saal, L. H. *et al.* Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nature Genet.* **40**, 102–107 (2008).
40. Saal, L. H. *et al.* Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc. Natl Acad. Sci. USA* **104**, 7564–7569 (2007).
41. Tynan, J. A., Wen, F., Muller, W. J. & Oshima, R. G. Ets2-dependent microenvironmental support of mouse mammary tumors. *Oncogene* **24**, 6870–6876 (2005).
42. Buggy, Y. *et al.* Ets2 transcription factor in normal and neoplastic human breast tissue. *Eur. J. Cancer* **42**, 485–491 (2006).
43. Park, E. S. *et al.* Heterologous tissue culture expression signature predicts human breast cancer prognosis. *PLoS One* **2**, e145 (2007).
44. Svensson, S. *et al.* ERK phosphorylation is linked to VEGFR2 expression and Ets-2 phosphorylation in breast cancer and is associated with tamoxifen treatment resistance and small tumours with good prognosis. *Oncogene* **24**, 4370–4379 (2005).
45. Pap, T. *et al.* Activation of synovial fibroblasts in rheumatoid arthritis: lack of Expression of the tumour suppressor PTEN at sites of invasive growth and destruction. *Arthritis Res.* **2**, 59–64 (2000).
46. White, E. S. *et al.* Negative regulation of myofibroblast differentiation by PTEN (phosphatase and tensin homolog deleted on chromosome 10). *Am. J. Respir. Crit. Care Med.* **173**, 112–121 (2006).
47. Gibson, G. E. & Huang, H. M. Oxidative processes in the brain and non-neuronal tissues as biomarkers of Alzheimer's disease. *Front. Biosci.* **7**, d1007–d1015 (2002).
48. Allred, D. C., Harvey, J. M., Berardo, M. & Clark, G. M. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod. Pathol.* **11**, 155–168 (1998).
49. Soule, H. D. & McGrath, C. M. A simplified method for passage and long-term growth of human mammary epithelial cells. *In Vitro Cell. Dev. Biol.* **22**, 6–12 (1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Rawahneh and J. Moffitt for histotechnical assistance, K. Kornacker, S. Cory and I. Vasudeva Murthy for bioinformatics assistance, P. Gulati for statistics assistance, the Ohio State University Human Tissue Resource Network and the Ohio State University Comprehensive Cancer Center Microarray, Nucleic Acids, Transgenics and Flow Cytometry Shared Facilities for technical assistance. *MMTV-ErbB2* mice were provided by W. Muller. This work was funded by the National Institutes of Health to G.L. (R01CA85619, R01HD47470, P01CA097189) and to M.C.O. (R01CA053271, P01CA097189), by the Komen Breast Cancer Foundation and Evelyn Simmers Charitable Trust to M.C.O., by the Terry Fox New Frontiers Group Grant to M.P., and by the Natural Science and Engineering Research Council of Canada Discovery Grants Program grant to M.H. F.L. and F.P. were funded by Department of Defense Pre-doctoral Fellowships and J.-L.C. was funded by a Department of Defense Postdoctoral Fellowship. G.L. is the recipient of the Pew Charitable Trusts Scholar Award and the Leukemia and Lymphoma Society Scholar Award. M.P. holds the Diane and Sal Guerrero Chair in Cancer Genetics at McGill University.

Author Contributions G.L. and M.C.O. designed and supervised this study, analysed data, and helped write and edit the manuscript. A.J.T., C.Z.C., F.L. and J.A.W. designed and performed experiments, collected and analysed data, and co-wrote the paper. N.C., J.C.T., H.W., J.-L.C., S.M.S. and M.N.G. technically assisted with experiments, and collected and analysed data. G.W., A.J.T., M.L.R. and M.W. performed experiments in initial stages of the project, particularly in designing and characterizing the mouse models. S.N., P.S. and T.J.R. contributed to the histopathological analysis of the mouse mammary tumour models and writing the manuscript. S.H.B. and L.Y. contributed to the histopathological analysis of human samples and writing the manuscript. S.A.F. and J.A.S. contributed to the statistical analyses of data and writing the manuscript. A.M., F.P., J.A.W., E.C., M.H. and M.P. contributed to the analysis and comparison of mouse and human microarray data and writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.C.O. (michael.ostrowski@osumc.edu) or G.L. (gustavo.leone@osumc.edu).

METHODS

Transgenic mice. Animals were housed according to federal and University Laboratory Animal Resources at Ohio State University regulations.

Fsp-cre: the generation of the *Fsp-cre* mouse line has been previously described¹⁵. *Ets2^{loxP}*: the *Ets2* conditional transgenic mouse line was generated such that the *Ets2* pointed domain is 'floxed'²⁸. The pointed domain is encoded by exons 3–5 and important for the protein–protein interaction and signal transduction.

Ets2^{db}: the *Ets2^{db}* mouse was a gift from R. G. Oshima³².

Pten^{loxP}: the *Pten^{loxP}* mice were generated in our laboratory. *LoxP* sites were introduced into two *HpaI* sites within introns 3 and 5 of the *Pten* gene, respectively, to flank exons 4 and 5. Exon 5 encodes the lipid phosphatase domain. Tissue-specific expression of *cre* will excise exons 4 and 5, generating a loss-of-function *Pten^d* allele (Supplementary Fig. 1).

Rosa-LacZ: the generation of the conditional *Rosa-LacZ* mouse line has been previously described⁵⁰. A list of PCR primers can be found in Supplementary Table 3.

Mammary tissue transplantation. Transplant procedure was based on a previously published method¹⁷. The day before surgery, recipient mice were anaesthetized with isoflurane (Abbott Laboratories) and a square area along the scapular region was shaved (roughly 25 mm × 25 mm). At the time of surgery, inguinal and groin mammary tissue (approximately 5 mm × 5 mm in size) was removed from 8- to 9-week-old donor females and placed subcutaneously into the scapular region of wild-type hosts (littermates or FVB/N purchased from Taconic) through two 5- to 10-mm incisions on the left and right side under aseptic conditions. Only tissue from the abdominal (#4 and #9) and groin (#5 and #10) mammary glands was used for transplant after the lymph nodes were removed. The small incisions were closed using a 9-mm wound clip. Animals were monitored twice a week until tumour onset. Mice were killed either at specific time points, when the tumour was about 2 cm in size, or it presented a health problem to the animal such as exterior ulceration at the site of the tumour.

Orthotopic mammary gland injection. The neu-expressing mouse mammary carcinoma cell line NT2.5 was provided by P. T. Kaumaya and was maintained as described³⁶. Eight-week-old female mice of each genotype were anaesthetized and injected with 5×10^5 NT2.5 cells at both inguinal mammary glands. Tumour initiation was monitored by palpating twice a week. All the mice were killed three weeks after injection. Tumour volume was calculated by the formula $V = \frac{1}{2} \times \text{length} \times (\text{width})^2$.

Tissue processing and X-gal staining. Large individual tumours (typically ~1–2 cm) or the remainder of transplanted tissues were removed, divided and either fixed in 4% paraformaldehyde for 24–48 h or embedded directly in OCT (Sakura). Fixed tissue samples were embedded in paraffin and cut into 5-μm sections for H&E, IHC or immunofluorescence staining. For each sample collected, two sets of sections were obtained at 25-μm intervals for analysis. Corresponding OCT-embedded tissue was sectioned (7 μm) in a similar manner for X-gal or immunofluorescence staining. For X-gal staining, frozen tissue sections were dried 15 min at room temperature before fixing in a glutaraldehyde solution (0.2% glutaraldehyde, 1.25 mM EGTA, pH 7.3 and 2 mM magnesium chloride in 1 × PBS) for 30 min. The sections were washed with LacZ wash buffer (2 mM magnesium chloride, 0.01% sodium deoxycholate, 0.02% IGEPAL CA-630 (Sigma) in PBS) for 5 min three times and then stained in LacZ staining solution (4 mM potassium ferricyanide, 4 mM potassium ferrocyanide, 1 mg ml⁻¹ X-gal in LacZ wash buffer) at 37 °C overnight (~18 h) and protected from light. Stained sections were washed in PBS for 5 min three times and then rinsed with water for 2 min before counter-staining with nuclear fast red. Collagen deposition was visualized using standard Masson's trichrome staining. Whole mammary gland sections were stained and imaged using an Aperio Scanscope CS whole-slide scanner. Areas of skin and muscle, which also stain positive, were manually encircled and excluded from the analysis.

IHC and immunofluorescence. IHC and immunofluorescence were performed using paraffin sections with the following antibodies: Pten (1:100, Cell Signaling, overnight), collagen I (1:100, Abcam, 30 min), F4/80 (1:50, Caltag, overnight), phospho-Akt^{S473} (1:50, Cell Signaling, overnight), phospho-JNK^{T183/Y185} (1:50, Cell Signaling, overnight), phospho-Erk1/2 (1:100, Cell Signaling, overnight), phospho-Ets2^{T772} (1:25, M.C.O. laboratory, overnight), Cytokeratin 8/18 (1:300; Research Diagnostics, 30 min), E-cadherin (1:700, BD-Pharmingen, 30 min), mouse α-SMA (1:200, Sigma, 30 min) and collagen IV (1:100, Chemicon, 30 min). In general, paraffin sections were deparaffinized and the antigen retrieval accomplished by incubation in antigen retrieval solution (DAKO) at >95 °C (30 min). For IHC, staining was developed using the biotin/avidin/horseradish peroxidase system from Vector Laboratories according to the manufacturer's instructions. All IHC slides were counterstained with haematoxylin and images obtained using an Eclipse 50i microscope (Nikon) and an AxioCam HRc camera (Zeiss). For IHC quantification, cells were counted manually from

ten random fields and reported as a percentage of positive cells from the total cell population.

Frozen sections of mammary glands for immunofluorescence were fixed at 4 °C in either 4% paraformaldehyde in PBS or acetone. Samples were treated with phospho-Ets2^{T772} (1:125; M.C.O. laboratory, overnight), vimentin (1:50; Santa Cruz Biotech, overnight), Pten (1:100, Cell Signaling, overnight), VEGF₁₆₄ (1:100, R&D Systems, overnight), phospho-VEGF2^{Y1173} (1:100, Cell Signaling, overnight) and CD31 (1:50, BD Biosciences, 30 min) primary antibodies. Fluorescent staining was developed using secondary antibodies conjugated to AlexaFluor dyes following standard protocol (Invitrogen; Molecular Probes). For Pten immunofluorescence on paraffin sections, the signal was amplified using a biotinylated secondary antibody and Texas red conjugated to streptavidin. All immunofluorescence sections were counterstained with DAPI and images obtained using an Axioscope 40 microscope (Zeiss) equipped with an AxioCam HRc camera (Zeiss). Image analysis for quantifying CD31 and phospho-VEGF2^{Y1173} areas was performed using standard image processing techniques implemented in Matlab. VEGF₁₆₄ images were quantified with ImageJ software.

Gelatinase *in situ* zymography. This was performed as described, with minor modification⁵¹. Briefly, frozen sections (10 μm) were quickly fixed with cold acetone, rehydrated with PBS and then incubated with 40 μg ml⁻¹ DQ-gelatin fluorescein conjugate (Molecular Probes) in reaction buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM CaCl₂ and 0.2 mM Na₂S₂O₈, pH 7.6) for 10 h. The reaction was quenched with 10 mM EDTA-PBS wash. Nuclei were counterstained with DAPI. A consecutive slide was stained with H&E to localize the MMP9 activity. Fluorescent images were acquired with Axioscope40 microscope (Zeiss) and AxioCam HRc camera (Zeiss).

Western blot. One million to two million cells were lysed with radio immunoprecipitation assay (RIPA) buffer (50 mM PH7.4 Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate and 0.1% SDS) containing protease and phosphatase inhibitors (Roche). Primary antibodies for Pten (Cell Signaling), Akt (Cell Signaling), phospho-Akt^{S473} and phospho-Akt^{T308} (Cell Signaling), phospho-JNK^{T183/Y185} (Cell Signaling), phospho-Erk1/2 (Cell Signaling), Erk1/2 (Santa Cruz), Ets2 (M.C.O. laboratory) and Collagen 1 (Abcam) were incubated overnight at a dilution of 1:1,000, whereas tubulin (Sigma) was used at 1:5,000 dilution. Washed membranes were blotted with either horseradish-peroxidase-conjugated anti mouse IgG or anti rabbit IgG antibodies and developed with enhanced chemiluminescent substrate (Thermo Scientific).

Microarray analysis. The data were analysed using WEDGE++ expression analysis⁵². Heat map representation (Supplementary Fig. 6) was performed using the TIGR Multiexperiment Viewer program MeV version 4.1.

Quantitative RT–PCR. Quantitative gene expression was performed using 50 ng complementary DNA per reaction. Taqman Roche Universal Probe Library system probe and primers (Roche) following the manufacturer's instructions. Reactions were performed on the I-cycler iQ Real-Time machine (Bio-Rad). A list of quantitative RT–PCR primers can be found in Supplementary Table 3. The reference gene used for all quantitative RT–PCR assays is *Rpl4*.

ChIP and quantitative RT–PCR. ChIP assays were performed as described by Hu *et al.*⁵³. Primary fibroblasts were cross-linked with 1% formaldehyde and soluble chromatin was prepared with sonication to an average DNA length of 200–1,000 base pairs. Sheared soluble chromatin was pre-cleared with transfer-RNA-blocked Protein G Agarose, and 10% of the pre-cleared chromatin was set aside as input control. Immunoprecipitation was performed with 5 μg of Ets2 antibody or rabbit IgG overnight at 4 °C. Immune complexes were pulled down with Protein G-Agarose, washed, and eluted with elution buffer (0.1 M NaHCO₃, 1% SDS), and cross-links removed by incubating with 200 mM NaCl containing 50 μg ml⁻¹ of RNase A (Sigma) at 65 °C overnight. DNA was purified with the Qiagen PCR Purification Kit after proteinase K treatment according to the manufacturer's instructions. Samples were analysed by quantitative RT–PCR as indicated above. The threshold for the promoter being studied was adjusted by that of input values and represented as relative abundance. All quantitative RT–PCR reactions were analysed by melt curve analysis and agarose gels to confirm the specificity of the reaction. A list of ChIP primers can be found in Supplementary Table 3.

Generating the human stroma heat map with genes identified in *Pten* null fibroblasts. Analysis of the *Pten* null mouse microarray data using WEDGE++ software⁵² led to the identification of 195 differentially expressed probe sets matching to 150 unique mouse genes. A search for human orthologues using Ensembl and MGI databases yielded a list of 137 genes. These genes were queried against the McGill Cancer Center's Breast Stroma Microarray data (GSE9014 and GSE4823). Of the 137 genes, 129 were represented on the Agilent Custom Array used in the McGill study. A heat map was generated for the human stroma data set (52 normal stroma and 49 tumour stroma samples). To achieve better resolution

on the heat map, and to identify only those genes that had highly variable gene expression across all samples, a variance cutoff of >0.5 was used to generate a subset of 70 genes. The heat map (Fig. 5a) shows the ability of these 70 genes to separate the normal and tumour stroma samples based solely on their gene expression profiles. This partitioning is highly significant ($P = 3.9 \times 10^{-15}$), as determined by Wilcoxon's test on the average expression of *Pten* null-signature in all samples.

Interestingly, comparison of 137 human orthologues of the *Pten* null list with the 163 genes associated with recurrence in tumour stroma from the McGill study showed that there are 12 genes (highlighted in red; Fig. 5a) present on both gene lists. This overlap is again highly significant ($P = 2.5 \times 10^{-8}$; Fisher's exact analysis).

TMA. TMA slides containing paraffin-embedded breast carcinoma tissues were processed at the Pathology Core Facility and Tissue Archives Human Tissue Resource Network at Ohio State University. IHC on these slides was performed with phospho-Ets2^{T72} (1:25, M.C.O. laboratory), Pten (1:100, Cell Signaling) and phospho-Akt^{S473} (1:50, Cell Signaling) following standard procedures. Level of PTEN and phospho-ETS2 and phospho-AKT expression in stromal cells were scored with Allred score method⁴⁸.

Statistical analysis. The number of animals and experiments are indicated in the figures. All the statistical tests were two-sided. The Wilcoxon rank test was used for some of the statistical analyses when the outcome variables were skewed and could not be transformed to achieve normality (Figs 1f and 2c, e, i). General linear models (analysis of variance) were used to study differences in continuous outcome variables among groups. *Ad hoc* pairwise comparisons were performed if the overall test was significant (Fig. 4b, d, f and Supplementary Figs 9c and 11b). Fisher's exact test was used to compare counts in Fig. 1e because low counts (below 5) were observed in one of the groups. Fisher's exact test was also used to compare gene overlap in Fig. 5b. Multiplicity adjustment methods (Bonferroni method) were used when multiple pairwise comparisons were performed. A two-sample (unpaired) Student's *t*-test was used for tumour load and the fluorescent microscopy images of VEGF₁₆₄, CD31 and pVEGFR^{Y1173} staining in Fig. 3b, e, f and Supplementary Fig. 9a, e. For human TMA, a Pearson correlation

was used to evaluate the association of Allred scores between PTEN/P-ETS2^{T72}, PTEN/P-AKT^{S473} and P-ETS2^{T72}/P-AKT^{S473}.

We show the expression data comparing normal and tumour stroma in Fig. 5a as a heat map because it highlights the 'direction' of the change in expression. Moreover, a permutation test strategy was used to study whether the 70-gene subset of the mouse *Pten* signature discriminated between human tumour and normal stroma samples⁵⁴ (Fig. 5a). A total of 100,000 random permuted samples was used to compute the empirical *P* value of the 70-gene signature. A Wilcoxon rank sum value for each random 70-gene permutation was computed from the average log expression difference (between normal and tumour stroma) of the 70 genes. For computing the average log expression difference of the mouse *Pten* 70-gene signature, we assigned negative values to the 13 downregulated genes. The empirical *P* value obtained using this method (8×10^{-5}) represents the proportion of random permuted samples that had a Wilcoxon rank sum value larger than the Wilcoxon rank sum value of the *Pten* 70-gene unpermuted signature. In addition, principal component analysis was used to explore whether the *Pten* 70-gene signature discriminated normal versus tumour stroma; this principal component analysis, like the heatmap presented in Fig. 5a, showed perfect separation of normal and tumour samples (data not shown), with the first component being sufficient for this separation. A Wilcoxon rank sum test yielded a *P* value of less than 1×10^{-10} .

50. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
51. Mook, O. R., Van Overbeek, C., Ackema, E. G., Van Maldegem, F. & Frederiks, W. M. *In situ* localization of gelatinolytic activity in the extracellular matrix of metastases of colon cancer in rat liver using quenched fluorogenic DQ-gelatin. *J. Histochem. Cytochem.* **51**, 821–829 (2003).
52. Auer, H. *et al.* Gene-resolution analysis of DNA copy number variation using oligonucleotide expression microarrays. *BMC Genomics* **8**, 111 (2007).
53. Hu, R. *et al.* Eos, MITF, and PU.1 recruit corepressors to osteoclast-specific genes in committed myeloid progenitors. *Mol. Cell. Biol.* **27**, 4018–4027 (2007).
54. Westall, P. H. & Young, S. S. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (John Wiley, 1993).

ARTICLES

SSB protein diffusion on single-stranded DNA stimulates RecA filament formation

Rahul Roy^{1,2,†}, Alexander G. Kozlov³, Timothy M. Lohman³ & Taekjip Ha^{1,2,4}

Single-stranded DNA generated in the cell during DNA metabolism is stabilized and protected by binding of ssDNA-binding (SSB) proteins. *Escherichia coli* SSB, a representative homotetrameric SSB, binds to ssDNA by wrapping the DNA using its four subunits. However, such a tightly wrapped, high-affinity protein–DNA complex still needs to be removed or repositioned quickly for unhindered action of other proteins. Here we show, using single-molecule two- and three-colour fluorescence resonance energy transfer, that tetrameric SSB can spontaneously migrate along ssDNA. Diffusional migration of SSB helps in the local displacement of SSB by an elongating RecA filament. SSB diffusion also melts short DNA hairpins transiently and stimulates RecA filament elongation on DNA with secondary structure. This observation of diffusional movement of a protein on ssDNA introduces a new model for how an SSB protein can be redistributed, while remaining tightly bound to ssDNA during recombination and repair processes.

The primary activity of SSB proteins in DNA metabolism is to bind preferentially to ssDNA with high affinity independently of sequence¹. However, SSB proteins also have a crucial role by interacting with a large number of proteins, directing these proteins to sites of DNA replication, recombination or repair². SSB proteins are often viewed as providing only inert protection for transiently formed ssDNA; however, there is increasing evidence that SSB–ssDNA complexes are highly dynamic, which can be functionally important^{3–5}. The *E. coli* SSB forms a stable homotetramer and can bind ssDNA in several modes with different properties⁶. In particular, under relatively high salt conditions (≥ 200 mM NaCl or ≥ 2 mM Mg^{2+} or polyamines), a low cooperativity complex forms in which ~ 65 nucleotides of ssDNA wrap around the tetramer ((SSB)₆₅ mode), interacting with all four subunits in such a way that the two ends of the ssDNA exit the protein in close proximity (referred to here as ‘closed’ wrapping)^{6,7}.

Owing to its transient role in replication, recombination and repair processes, SSB must be recycled (dissociate and reassociate with ssDNA) as well as repositioned within its ssDNA complexes. However, because *E. coli* SSB binds with extremely high affinity to ssDNA making several binding interactions⁶, it remains unclear how SSB is displaced rapidly by other proteins, for example DNA polymerase or RecA, for subsequent DNA processing. Here, we demonstrate that an *E. coli* SSB tetramer can migrate via a random walk along ssDNA, thus providing a mechanism by which it can be repositioned along ssDNA while remaining tightly bound.

Diffusional migration of SSB on ssDNA

To investigate potential movements of SSB on ssDNA, we used single-molecule fluorescence resonance energy transfer (smFRET)^{8,9}. FRET efficiencies E from individual immobilized partial duplex DNA with a 3' (dT)_{*n*} tail ($64 \leq n \leq 131$) bound to SSB were acquired using total internal reflection fluorescence microscopy⁹. Surface immobilization and fluorescent labelling have no measurable effect on the dynamics of SSB binding mode transitions⁵. Owing to the closed wrapping in the (SSB)₆₅ binding mode favoured under our conditions (500 mM NaCl

or 10 mM Mg^{2+})⁷, when SSB is bound to ssDNA of 65–70 nucleotides with its two ends labelled with donor (Cy3) and acceptor (Cy5) fluorophores, singular high FRET distributions were observed⁵. However, when a (dT)₆₉ tail is further extended by an additional 12 nucleotides of sequence complementary to the overhanging cohesive end of l-strand of λ -phage DNA, individual SSB–ssDNA complexes display large FRET fluctuations in the millisecond timescale (Fig. 1a). These

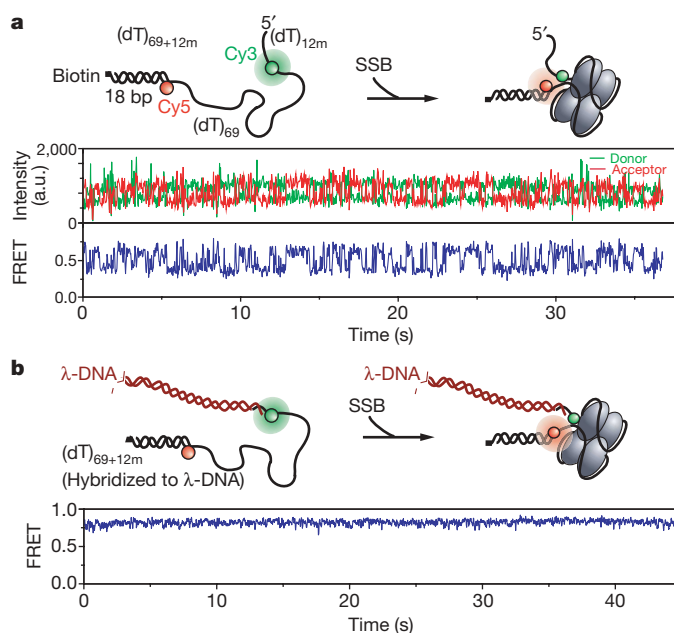


Figure 1 | FRET fluctuations arising from diffusional migration of SSB on ssDNA. **a**, With the 12-nucleotide extension to the (dT)₆₉ separating the donor and acceptor fluorophores, rapid fluctuations between several FRET states are observed due to diffusion of SSB on the ssDNA. a.u., arbitrary units. **b**, When the 12-nucleotide extension is hybridized to a complementary sequence, only steady high FRET values are observed.

¹Center for Biophysics and Computational Biology, ²Department of Physics and Center for the Physics of Living Cells, University of Illinois, Urbana-Champaign, Illinois 61801, USA. ³Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴Howard Hughes Medical Institute, Urbana, Illinois 61801, USA. [†]Present address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

fluctuations were markedly suppressed when the 12-nucleotide extension is hybridized to a cohesive end of a λ -DNA (Fig. 1b). To exclude binding and dissociation of additional SSB molecules as the cause of fluctuations, unbound SSB was removed by a buffer wash before measurements. DNA unwrapping/rewrapping dynamics, occurring in tens of microseconds in high salt concentration^{3,4}, is completely averaged out within our 10–30-ms time resolution⁵. We also ruled out local melting of the duplex portion as a source of fluctuation (Supplementary Information). Therefore, these fluctuations must arise from additional conformational states enabled by the 12-nucleotide extension.

To test whether the FRET fluctuations are caused by transient excursions of SSB to the extension, we varied the length of the extension ((dT)_m, $n = 0$ –18) while keeping the ssDNA between Cy3 and Cy5 at 69 or 70 nucleotides. If an SSB tetramer binds randomly and remains fixed at the initial site of binding undergoing only transient interactions with ssDNA outside the binding site, each complex will generate a FRET distribution that is unique to the initial site of binding. However, all complexes for each construct displayed similar FRET time trajectories (Supplementary Fig. 1). Furthermore, if SSB migrates along the DNA, larger excursions away from the high FRET state are expected for longer extensions. Indeed, average FRET values decreased for longer extensions while the high FRET state was still transiently visited (Supplementary Fig. 1). The FRET distribution and the timescale of fluctuations are relatively independent of the salt concentration (Supplementary Fig. 2), presenting evidence against these FRET changes arising from binding mode transitions that display a strong salt dependence^{10–12}. Hence, these fluctuations probably reflect SSB's diffusional migration on ssDNA with the different FRET values corresponding to different SSB locations.

To make unbiased assignments of FRET states, we used a hidden Markov model (HMM)-based statistical approach that determines the most likely time sequence of FRET states (Fig. 2a)^{13,14}. The result is further reduced to a transition density plot^{13,14} that allows the number of distinct FRET states, their FRET values and the transition rates to be estimated (Fig. 2b). We analysed SSB migration on DNA molecules with several 3' dT tail lengths (0–12 nucleotide extension beyond 65 nucleotide binding site size) at 13 °C to slow down migration (Fig. 2a, b). Longer extensions gave several indistinguishable low FRET states in the transition density plot (Supplementary Fig. 3). For (dT)₆₉₊₈ (12 nucleotide extension from the 65 nucleotide binding site size with 69 nucleotide separation between fluorophores), six distinct FRET states were resolved (Fig. 2b) with transitions occurring between

nearest neighbours. We assigned the highest FRET value ($E \approx 0.8$) to the state with SSB closest to the single-stranded–double-stranded DNA junction and lower FRET values for positions away from the junction. The rates of transition, or the 'stepping rates', were independent of the beginning and ending state of transition (Supplementary Fig. 4) and ranged between 3.0 and 4.5 s⁻¹ (Fig. 2c). Similar analysis yielded 5, 3 and 2 states for DNA with 8-, 2- and 0-nucleotide extensions, respectively (Supplementary Fig. 3). Therefore, every 2–4 nucleotides of DNA extension provides an additional configuration, yielding an apparent step size of about 3 nucleotides.

Because FRET fluctuations became too fast for HMM analysis above 13 °C, we used autocorrelation analysis of FRET efficiency E for the temperature dependence studies (Fig. 2d). The averaged autocorrelation function plots of the SSB-(dT)₆₉₊₈ complexes were best fit by bi-exponential decays. The shorter lifetime was equal to the time resolution independent of temperature and is ascribed to photophysical or detection noise. The longer lifetime, τ_{long} , displayed a monotonic temperature dependence and was attributed to SSB diffusion. The Arrhenius fit of $\ln(1/\tau_{\text{long}})$ against $1/T$ (Fig. 2e) gave an apparent activation energy of 81 ± 7 kJ mol⁻¹. Combined with the stepping rate of ~ 4 s⁻¹ at 13 °C, we can then estimate a stepping rate of ~ 60 s⁻¹ at 37 °C. Assuming a 3-nucleotide step size, the diffusion coefficient of an SSB tetramer along ssDNA at 37 °C is estimated to be 270 (nucleotides)² s⁻¹.

As a further test of SSB migration on the ssDNA, we used single-molecule three-colour FRET^{9,15} using a donor-labelled SSB mutant (A122C labelled with \sim one Alexa555 per SSB tetramer) and two different acceptors, Cy5 and Cy5.5, attached to the two ends of a (dT)₁₃₀ (Fig. 3a). The large separation between the two acceptors eliminates any significant FRET between them. If a single SSB tetramer diffuses on the long ssDNA, high FRET events to either acceptor will be mutually exclusive. Indeed, we observed rapid and anti-correlated fluctuations of apparent FRET efficiencies to the two acceptors, $E_{\text{app},5}$ and $E_{\text{app},5.5}$, demonstrating that SSB truly diffuses on the DNA (Fig. 3a, b). To ensure only single SSB molecules were bound on the DNA, 1 min incubation with sub-saturating concentrations of SSB (<100 pM) was followed immediately with a buffer wash and only traces displaying single donor photobleaching events were analysed. At higher SSB concentration (10 nM), much slower FRET fluctuations were observed, probably due to binding of additional SSB (Supplementary Fig. 5).

To probe how far SSB can move on a long ssDNA, we placed Cy5 and Cy5.5 on the two ends of a (dT)₁₃₀ and Cy3 in the middle (named

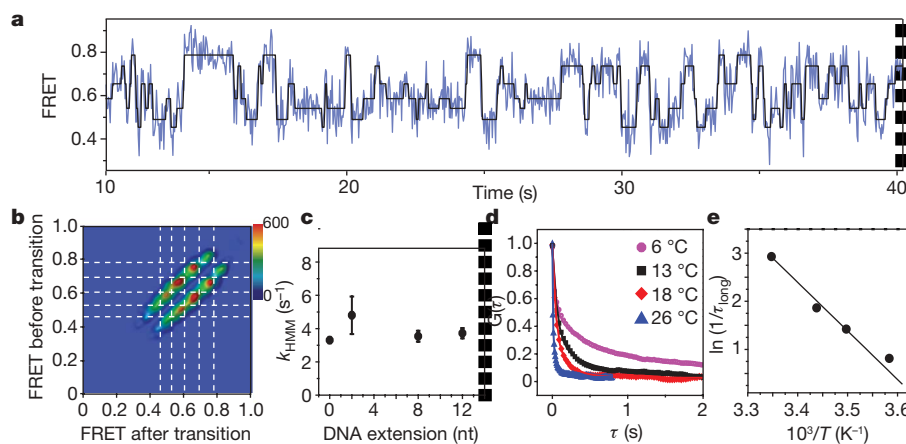


Figure 2 | Analysis of SSB mobility on ssDNA. **a**, Hidden Markov model (HMM)-derived idealized FRET trajectory (black) superimposed on the FRET trajectory (blue) of a single SSB-(dT)₆₉₊₈ complex. **b**, Transition density plot for (dT)₆₉₊₈ DNA. **c**, Average rates of SSB migration as a function of DNA extension length from HMM analysis. Error bars are

standard error over $2(m - 1)$ values of the transition rate obtained from HMM analysis where m is the number of distinct FRET states ranging from 2 to 6. **d**, Autocorrelation ($G(\tau)$) analysis of FRET trajectories for (dT)₆₉₊₈ DNA fit to bi-exponential decay function for different temperatures (T). **e**, Arrhenius plot of apparent rates as a function of $1/T$.

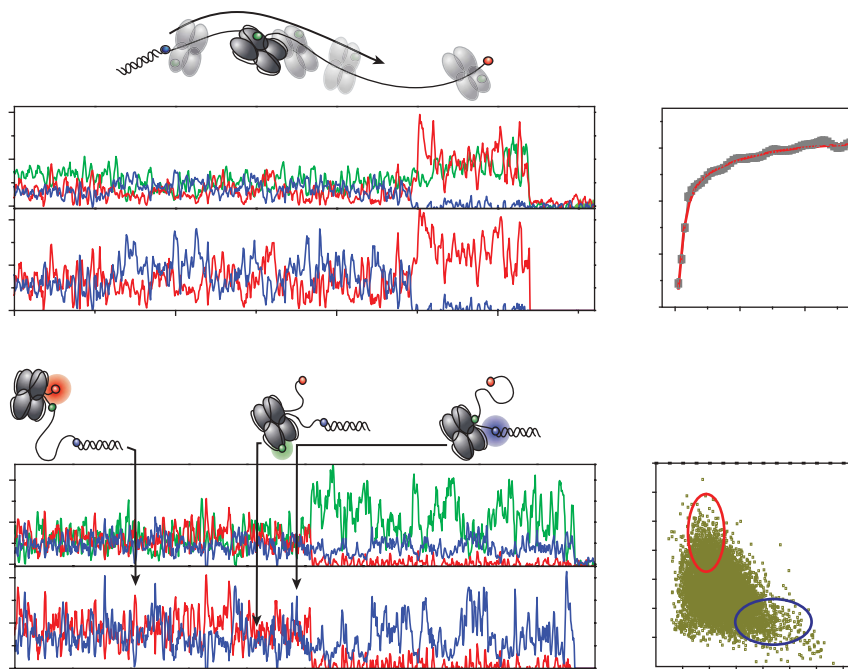


Figure 3 | SSB diffusion on ssDNA probed with three-colour FRET. **a**, Time traces of three colour intensities and the FRET efficiencies ($E_{app,5}$, $E_{app,5.5}$) display diffusion of donor-labelled SSB to acceptor-labelled ends of a $(dT)_{130}$ DNA. **b**, The average cross-correlation $XC(\tau)$ between $E_{app,5}$ and $E_{app,5.5}$ time traces (48 molecules) fit with a bi-exponential decay (red), demonstrating

anti-correlated fluctuations. **c**, A $(dT)_{130}$ with a centrally placed donor, and acceptors at the two ends, displays excursions of unlabelled SSB to either extremity of the DNA resulting in high FRET for the corresponding acceptor. **d**, Scatter plot of $E_{app,5.5}$ against $E_{app,5}$ values (35 molecules) show mutually exclusive high FRET events (oval regions).

$(dT)_{65+65}$). This three-colour FRET scheme allows us to determine at which end the SSB was present by following the ‘closed’ wrapping of that DNA segment and high FRET to the corresponding acceptor (Fig. 3c). Both the dye pairs display transient high FRET states that are anti-correlated, indicating that the same SSB molecule was capable of migrating to either end of the DNA (Fig. 3c, d). Therefore, SSB can move at least 65 nucleotides via diffusion and is not constrained to its initial binding site.

SSB displacement by RecA filament

SSB modulates the interaction between the RecA protein and ssDNA in the SOS response and recombinational repair pathway^{2,16–19} and mutations in the *ssb* gene cause inefficient recombinational repair and homologous recombination^{1,20–22}. A RecA filament can readily displace SSB from the DNA if assisted by RecFOR, χ -modified RecBCD or a preassembled nucleation cluster^{14,23–27}. However, the mechanism of efficient SSB displacement by RecA was unclear given the tight binding of SSB to ssDNA.

Our estimated diffusion step size of SSB, ~ 3 nucleotides, is the same as the binding site size of a RecA monomer which is the unit of filament extension^{14,28}. We therefore hypothesized that a monomer-by-monomer addition of RecA to the DNA segment freed up by SSB diffusion might convert the random walk of SSB into unidirectional movement (Supplementary Movie 1). To test this idea, we devised a three-colour FRET assay using a DNA with a 96 nucleotide 3' tail, $(dT)_{30+65}$, labelled at positions 0, 30 and 95 with Cy5.5, Cy3 and Cy5, respectively (Fig. 4). The apparent FRET efficiencies of DNA only are low for both acceptors (~ 0.1), and drop to zero on RecA-ATP γ S filament formation (Fig. 4a, b). SSB addition after flushing out excess RecA and ATP γ S removes the RecA-ATP γ S filament from the ssDNA tail, but not from the duplex DNA¹⁴, and the ssDNA wraps around SSB, displaying higher FRET with a broad distribution that reflects SSB diffusion (Fig. 4c). The RecA-ATP γ S filament remaining on the duplex serves as the nucleation cluster for filament elongation on the 3' ssDNA tail¹⁴ such that on addition of RecA and ATP, the elongating filament rapidly replaces SSB on the ssDNA ($E_{app} = 0$) (Fig. 4e).

Figure 4f shows the real time three-colour FRET trajectories of SSB displacement by an elongating RecA filament. Before elongation, the FRET values fluctuate rapidly due to SSB diffusion. On addition of RecA and ATP, $E_{app,5.5}$ drops first as a RecA-ATP filament initiates at the ssDNA–dsDNA junction. As this filament grows further, $E_{app,5}$ attains a steady high value. Because the Cy3 and Cy5 are separated by 65 nucleotides at the distal DNA end, we attribute this increase in $E_{app,5}$ to the repositioning of SSB to the distal end, pushed by the elongating RecA filament. Finally, complete filament elongation gives $E_{app} \approx 0$, probably accompanied by SSB dissociation. Direct excitation of the acceptors afterwards confirmed that they were not photobleached.

We used exponential fits of average FRET curves to estimate the rates of three distinct events following the addition of RecA and ATP (Fig. 4g). (1) $E_{app,5.5}$ drops from 0.3 to 0 at a rate of $k_{1,SSB} = 0.24 \pm 0.02 \text{ s}^{-1}$. We assign this to RecA filament initiation from the RecA-ATP γ S nucleation site because once initiated, the decrease in $E_{app,5.5}$ is nearly instantaneous. (2) $E_{app,5}$ increases from 0.3 to 0.75 at the rate of $k_{2,SSB} = 0.2 \pm 0.01 \text{ s}^{-1}$. We assign this to RecA filament initiation and elongation by approximately ten RecA monomers on 30 nucleotides and SSB movement to the distal DNA end. The time intervals between the drop in the $E_{app,5.5}$ and the rise of $E_{app,5}$ place a lower limit of $\sim 0.6 \text{ s}^{-1}$ for the RecA elongation on a 30-mer of SSB-bound ssDNA (Supplementary Fig. 6). (3) The decrease of $E_{app,5}$ (traces synchronized when the high $E_{app,5}$ state is obtained) that we assign to SSB dissociation occurs at a much lower rate of $0.07 \pm 0.01 \text{ s}^{-1}$. The rates of filament initiation and elongation without SSB are comparable to those obtained with SSB (Fig. 4h and Supplementary Information). Furthermore, the rate of RecA elongation on bare ssDNA is about 20 s^{-1} per monomer at $1 \mu\text{M}$ RecA¹⁴ and is similar to the lower limit we determined here with SSB ($\sim 6 \text{ s}^{-1}$ per monomer), indicating that any hindrance to RecA elongation by SSB is minimal. Similar rates were observed on longer DNA where up to two SSB tetramers can bind (Supplementary Fig. 7).

Overall, the rate of SSB removal from the DNA end is approximately tenfold slower than what is expected from filament elongation alone. This observation indicates that SSB diffusion is important for

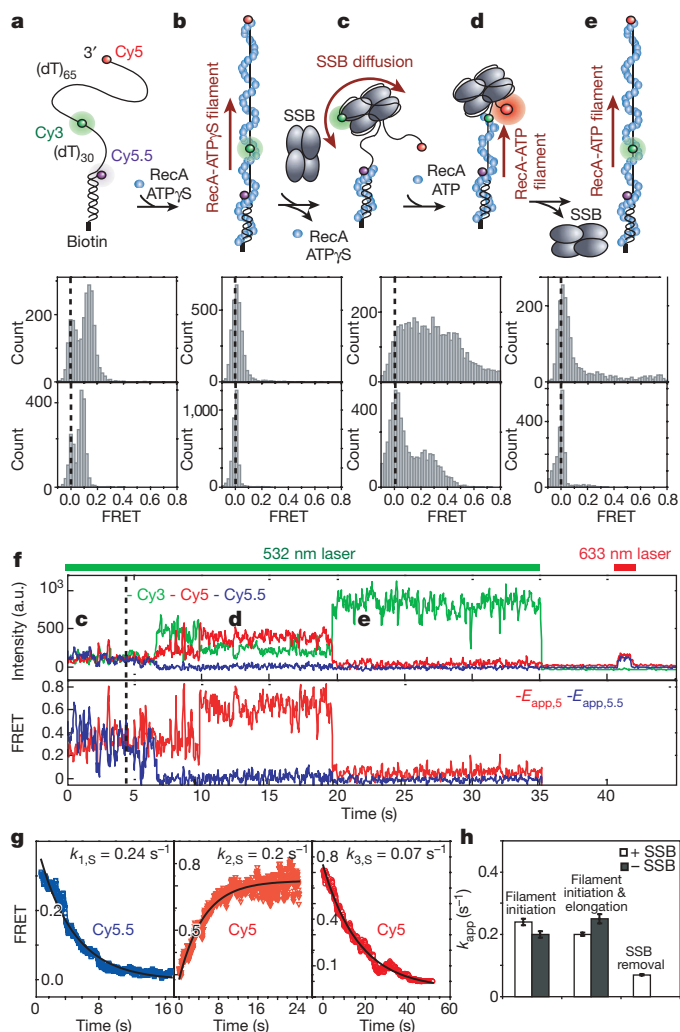


Figure 4 | Mechanism of SSB displacement by an extending RecA filament. **a–e**, Top panel, schematic of reaction steps; middle panel, $E_{app,5}$ histograms; bottom panel, $E_{app,5.5}$ histograms. **a**, DNA construct with Cy3, Cy5 and Cy5.5. **b**, RecA-ATP γ S filament (5 min incubation). **c**, SSB displaces RecA filament (15 min incubation). **d**, RecA filament growth. **e**, Filament completion (2 min incubation). **f**, Three-colour FRET trajectories for segments **c**, **d** and **e**. Excitation (633 nm) at 41 s confirms active acceptors. **g**, Sub-reaction kinetics with exponential fits ($n = 46$ molecules). Left panel, average $E_{app,5.5}$ decay; middle panel, average $E_{app,5}$ increase; right panel, average $E_{app,5}$ decay after maximum. **h**, The rates of RecA filament initiation and elongation, and of SSB removal. Error bars are propagated standard errors from exponential fits in **g**.

RecA filament elongation on SSB-coated DNA. This is because before SSB reaches the DNA end, its diffusion is isoenergetic and therefore rapid, whereas its further diffusion at the 3' end is energetically costly. This model of rectifying SSB diffusion by the directional growth of a RecA filament does not require any direct interaction of the two proteins²⁹ and hence could provide a general mechanism for displacement of SSB by proteins moving directionally on the ssDNA.

SSB diffusion aids RecA on DNA hairpin

SSB inhibits RecA filament formation at low salt and high SSB concentrations^{23,29}, but stimulates RecA filament formation in high salt concentration²⁹, probably by disrupting DNA secondary structures^{30,31}. Tetrameric SSBs can destabilize a DNA duplex possessing a single-strand tail that is shorter than the SSB binding site size³², but no significant duplex disruption was observed for a tail length equal to or greater than the binding site size (Fig. 1b and Supplementary Information). We therefore investigated whether SSB can disrupt a physiologically more relevant structure, that is, a hairpin flanked by

two single-stranded regions. The melting of a hairpin with a 7-base-pair stem and 3-nucleotide loop, *hp*, (Fig. 5a) is monitored via FRET between Cy3 and Cy5 attached to the ends of the hairpin in two different constructs, (dT)₆₅+*hp*+3 and (dT)₆₅+*hp*+65. A single high FRET population for an intact hairpin is partially replaced by lower FRET populations with SSB, signifying different states of hairpin unzipping (Fig. 5b). Single-molecule trajectories showed unzipping of the hairpin (Fig. 5c) with a majority displaying two-step unzipping with rate constants of $\sim 1.1\text{--}1.5\text{ s}^{-1}$ (Fig. 5c, d; details in Supplementary Information). Hence, a single SSB tetramer transiently disrupts DNA secondary structures as stable as a 7-base-pair stem by repositioning itself on and off the hairpin segment.

Finally, we tested whether such transient melting of a DNA hairpin by SSB promotes RecA filament formation on the hairpin. Starting from a pre-nucleated RecA-ATP γ S complex, a RecA-ATP filament was formed on (dT)₆₅+*hp*+65 DNA (Fig. 5e) giving rise to a $E_{app} \approx 0$ population representing filament formation over the melted hairpin (Supplementary Information). Notably, filament formation over the hairpin occurred 40-fold faster when SSB was present, demonstrating that SSB stimulates filament elongation over ssDNA that can form stable secondary structures (Fig. 5f and Supplementary Information). Interestingly, for our second construct (dT)₆₅+*hp*+3, filament formation over the hairpin remained slow even with SSB (Fig. 5f). This dependence on hairpin position further indicates that transient hairpin disruption by SSB is necessary for efficient filament elongation for the following reason. RecA filament elongation towards the 3' end decreases the length of ssDNA available for SSB binding, forcing it to eventually dissociate. However, SSB dissociation occurs before the filament elongates to the hairpin region of (dT)₆₅+*hp*+3 such that SSB-induced hairpin melting is reversed before filament growth over the hairpin segment (Supplementary Information).

On the basis of these results, we propose that SSB diffusion along ssDNA in the low cooperative (SSB)₆₅ mode, where ssDNA is populated mostly with single or two tetramers³³, stimulates RecA filament elongation by transiently removing DNA secondary structures ahead of the filament, and that filament elongation via RecA monomer addition in turn directionally biases SSB diffusion (Fig. 5g and Supplementary Movie 1). For long ssDNA bound by several SSB tetramers, directional migration of an SSB tetramer caused by RecA filament elongation may increase the local SSB concentration and promote transitions to other binding modes from which SSB dissociation may be much more rapid^{3,5}. If so, the findings made here may also be relevant for the removal of several SSBs from longer ssDNA.

Mechanism and functions of SSB diffusion

How does SSB diffuse on ssDNA? One possibility is the previously suggested rolling mechanism^{4,34}. SSB rolling would occur via partial unwrapping of one end segment of ssDNA from an SSB tetramer followed by re-wrapping of the other end in its place (Supplementary Fig. 8 and Supplementary Movie 1), resulting in one-dimensional random walk of SSB on ssDNA. Although our results are consistent with the rolling model, a definitive conclusion awaits further investigations.

Our work represents the first demonstration of any protein diffusing on ssDNA. By facilitating the redistribution of a tightly bound SSB tetramer along the ssDNA without full dissociation, SSB diffusion may be used in a variety of cellular processes, for example, stabilization of specific denaturation sites on superhelical DNA^{35,36} and facilitation of primase activity by positioning the SSB on G4-phage-type priming systems³⁷. The C-terminal region of *E. coli* SSB interacts with a variety of DNA repair enzymes and facilitates localization of these enzymes in the vicinity of ssDNA^{38,39}, raising the possibility that SSB acts as a mobile platform on the ssDNA for the repair and recombination machinery. The presence of homologous SSB proteins even in metazoans suggests that a similar diffusion mechanism might operate over a wide range of species⁴⁰.

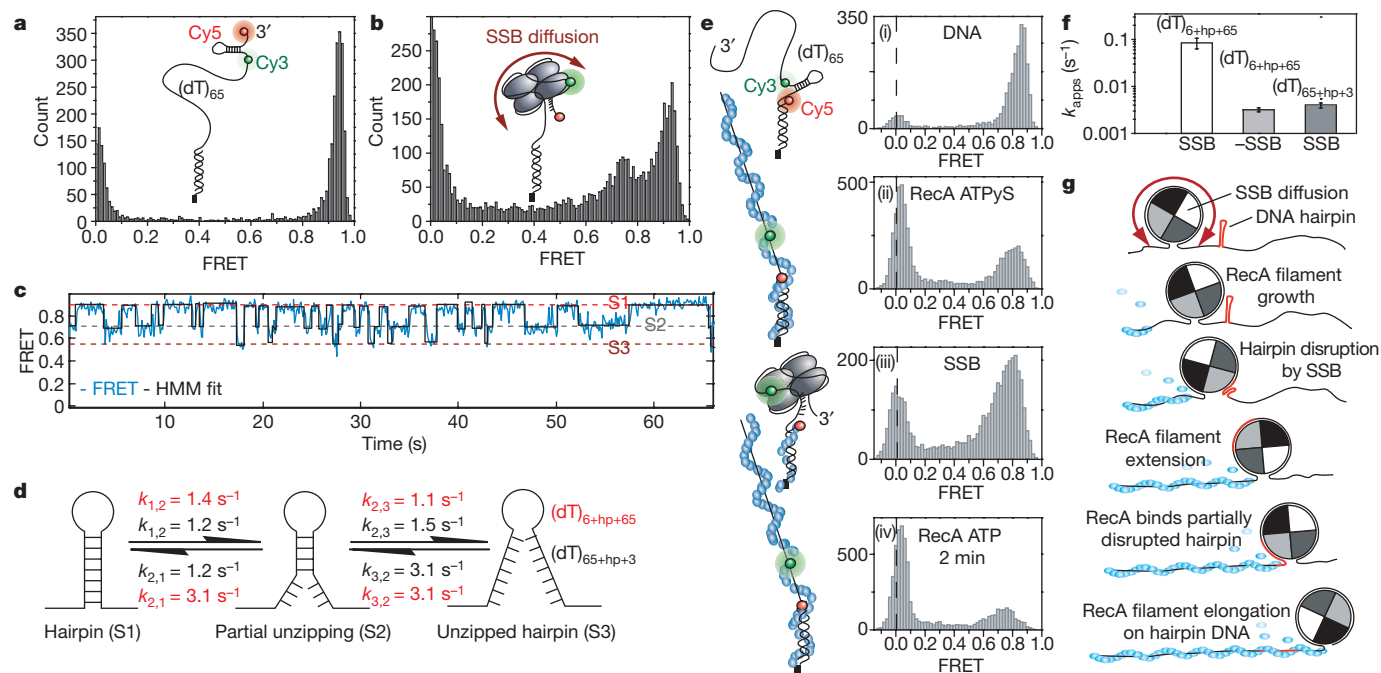


Figure 5 | SSB diffusion promotes RecA filament growth on DNA hairpin.

a, FRET histograms for $(dT)_{65+hp+3}$. **b**, Hairpin destabilization by SSB induces lower FRET states. **c**, FRET trajectory of $(dT)_{65+hp+3}$ shows fluctuations between states S1 (intact hairpin), and S2 and S3 (unzipped hairpin states); HMM-derived idealized trajectory (black). **d**, Transition rates between S1, S2 and S3 for $(dT)_{65+hp+3}$ and $(dT)_{65+hp+65}$. **e**, SSB-assisted RecA filament formation on hairpin DNA. (i) Intact hairpin;

(ii) RecA-ATP γ S filament formation on a majority of DNA; (iii) SSB replaces the RecA-ATP γ S filament restoring high FRET; (iv) RecA and ATP removes the hairpin structure. **f**, Rates of hairpin removal by extending RecA-ATP filament against hairpin position. Error bars are standard errors propagated from exponential fits (Supplementary Information). **g**, Model of SSB-assisted RecA filament growth on hairpin DNA.

METHODS SUMMARY

Partial duplex DNA (18 base pairs dsDNA) with 3' $(dT)_n$ tails (n ranging from 64 to 131 nucleotides) carrying one donor (Cy3) and up to two acceptors (Cy5 for two-colour FRET, Cy5 and Cy5.5 for three-colour FRET) were immobilized at the duplex end on a polyethylene glycol-coated surface using biotin-neutravidin and incubated with 100 pM–1 nM SSB in imaging buffer for 1 min before flushing. Single-molecule data were acquired using wide-field total-internal-reflection (TIR) fluorescence microscopy⁹ with 8–100 ms time resolution. All single-molecule measurements were performed at $23 \pm 1^\circ\text{C}$ unless specified otherwise in imaging buffer (10 mM Tris (pH 8.0), 500 mM NaCl, 0.1 mM Na_3EDTA , 0.1 mg ml^{-1} BSA) with oxygen scavenging system (0.5% w/v glucose, 1.5 mM Trolox⁴¹ or 1% β -mercaptoethanol, 165 U ml^{-1} glucose oxidase and 2,170 U ml^{-1} catalase). RecA-SSB experiments were conducted in 1 μM RecA (or 10 nM SSB), 1 mM ATP (or 1 mM ATP γ S) in 25 mM Tris acetate (pH 7.5), 50 mM sodium acetate, 10 mM magnesium acetate and 0.1 mg ml^{-1} BSA in combination with the oxygen scavenging system.

Received 19 March; accepted 20 August 2009.

Published online 11 October 2009.

- Meyer, R. R. & Laine, P. S. The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol. Rev.* **54**, 342–380 (1990).
- Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318 (2008).
- Kozlov, A. G. & Lohman, T. M. Kinetic mechanism of direct transfer of *Escherichia coli* SSB tetramers between single-stranded DNA molecules. *Biochemistry* **41**, 11611–11627 (2002).
- Kuznetsov, S. V., Kozlov, A. G., Lohman, T. M. & Ansari, A. Microsecond dynamics of protein–DNA interactions: direct observation of the wrapping/unwrapping kinetics of single-stranded DNA around the *E. coli* SSB tetramer. *J. Mol. Biol.* **359**, 55–65 (2006).
- Roy, R., Kozlov, A. G., Lohman, T. M. & Ha, T. Dynamic structural rearrangements between DNA binding modes of *E. coli* SSB protein. *J. Mol. Biol.* **369**, 1244–1257 (2007).
- Lohman, T. M. & Ferrari, M. E. *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.* **63**, 527–570 (1994).
- Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature Struct. Biol.* **7**, 648–652 (2000).

- Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93**, 6264–6268 (1996).
- Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nature Methods* **5**, 507–516 (2008).
- Bujalowski, W. & Lohman, T. M. *Escherichia coli* single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry* **25**, 7799–7802 (1986).
- Lohman, T. M. & Overman, L. B. Two binding modes in *Escherichia coli* single strand binding protein–single stranded DNA complexes. Modulation by NaCl concentration. *J. Biol. Chem.* **260**, 3594–3603 (1985).
- Griffith, J. D., Harris, L. D. & Register, J. III. Visualization of SSB–ssDNA complexes active in the assembly of stable RecA–DNA filaments. *Cold Spring Harb. Symp. Quant. Biol.* **49**, 553–559 (1984).
- McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951 (2006).
- Joo, C. *et al.* Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515–527 (2006).
- Hohng, S., Joo, C. & Ha, T. Single-molecule three-color FRET. *Biophys. J.* **87**, 1328–1337 (2004).
- Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
- Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehauer, W. M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58**, 401–465 (1994).
- Roca, A. I. & Cox, M. M. RecA protein: structure, function, and role in recombinational DNA repair. *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 129–223 (1997).
- Kuzminov, A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage λ . *Microbiol. Mol. Biol. Rev.* **63**, 751–813 (1999).
- Ennis, D. G., Amundsen, S. K. & Smith, G. R. Genetic functions promoting homologous recombination in *Escherichia coli*: a study of inversions in phage λ . *Genetics* **115**, 11–24 (1987).
- Glassberg, J., Meyer, R. R. & Kornberg, A. Mutant single-strand binding protein of *Escherichia coli*: genetic and physiological characterization. *J. Bacteriol.* **140**, 14–19 (1979).
- Golub, E. I. & Low, K. B. Indirect stimulation of genetic recombination. *Proc. Natl Acad. Sci. USA* **80**, 1401–1405 (1983).
- Umez, K., Chi, N. W. & Kolodner, R. D. Biochemical interaction of the *Escherichia coli* RecF, RecO, and RecR proteins with RecA protein and single-stranded DNA binding protein. *Proc. Natl Acad. Sci. USA* **90**, 3875–3879 (1993).
- Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ -regulated manner. *Cell* **90**, 77–86 (1997).

25. Bork, J. M., Cox, M. M. & Inman, R. B. The RecOR proteins modulate RecA protein function at 5' ends of single-stranded DNA. *EMBO J.* **20**, 7313–7322 (2001).
26. Morimatsu, K. & Kowalczykowski, S. C. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol. Cell* **11**, 1337–1347 (2003).
27. Hobbs, M. D., Sakai, A. & Cox, M. M. SSB protein limits RecOR binding onto single-stranded DNA. *J. Biol. Chem.* **282**, 11058–11067 (2007).
28. Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* **453**, 489–494 (2008).
29. Kowalczykowski, S. C., Clow, J., Somani, R. & Varghese, A. Effects of the *Escherichia coli* SSB protein on the binding of *Escherichia coli* RecA protein to single-stranded DNA. Demonstration of competitive binding and the lack of a specific protein–protein interaction. *J. Mol. Biol.* **193**, 81–95 (1987).
30. Kowalczykowski, S. C. & Krupp, R. A. Effects of *Escherichia coli* SSB protein on the single-stranded DNA-dependent ATPase activity of *Escherichia coli* RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J. Mol. Biol.* **193**, 97–113 (1987).
31. Muniyappa, K., Shaner, S. L., Tsang, S. S. & Radding, C. M. Mechanism of the concerted action of recA protein and helix-destabilizing proteins in homologous recombination. *Proc. Natl Acad. Sci. USA* **81**, 2757–2761 (1984).
32. Eggington, J. M., Kozlov, A. G., Cox, M. M. & Lohman, T. M. Polar destabilization of DNA duplexes with single-stranded overhangs by the *Deinococcus radiodurans* SSB protein. *Biochemistry* **45**, 14490–14502 (2006).
33. Bujalowski, W. & Lohman, T. M. Limited co-operativity in protein–nucleic acid interactions. A thermodynamic model for the interactions of *Escherichia coli* single strand binding protein with single-stranded nucleic acids in the “beaded”, (SSB)₆₅ mode. *J. Mol. Biol.* **195**, 897–907 (1987).
34. Römer, R., Schomburg, U., Krauss, G. & Maass, G. *Escherichia coli* single-stranded DNA binding protein is mobile on DNA: proton NMR study of its interaction with oligo- and polynucleotides. *Biochemistry* **23**, 6132–6137 (1984).
35. Clendenning, J. B. & Schurr, J. M. A model for the binding of *E. coli* single-strand binding protein to supercoiled DNA. *Biophys. Chem.* **52**, 227–249 (1994).
36. Glikin, G. C., Gargiulo, G., Rena-Descalzi, L. & Worcel, A. *Escherichia coli* single-strand binding protein stabilizes specific denatured sites in superhelical DNA. *Nature* **303**, 770–774 (1983).
37. Sun, W. & Godson, G. N. Structure of the *Escherichia coli* primase/single-strand DNA-binding protein/phage G4oric complex required for primer RNA synthesis. *J. Mol. Biol.* **276**, 689–703 (1998).
38. Shereda, R. D., Bernstein, D. A. & Keck, J. L. A central role for SSB in *Escherichia coli* RecQ DNA helicase function. *J. Biol. Chem.* **282**, 19247–19258 (2007).
39. Lecointe, F. et al. Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks. *EMBO J.* **26**, 4239–4251 (2007).
40. Richard, D. J. et al. Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* **453**, 677–681 (2008).
41. Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods* **3**, 891–893 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Joo, S. A. McKinney, I. Rasnik, S. Hohng and S. Myong for experimental help and discussion; C. Murphy, M. Nahas and K. Raghunathan for discussion; T. Ho and A. Niedziela-Majka for help with DNA and protein preparation, respectively; and R. Porter for the SSB expression plasmid. T.H. is an employee of the Howard Hughes Medical Institute. These studies were supported by grants from the National Institutes of Health and the National Science Foundation.

Author Contributions R.R., A.G.K., T.M.L. and T.H. designed the experiments, A.G.K. prepared the wild-type SSB protein and the mutant SSB with fluorescent labels, R.R. performed the experiments and analysed the data; R.R., T.M.L. and T.H. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.H. (tjha@illinois.edu).

LETTERS

Saturn's largest ring

Anne J. Verbiscer¹, Michael F. Skrutskie¹ & Douglas P. Hamilton²

Most planetary rings in the Solar System lie within a few radii of their host body, because at these distances gravitational accelerations inhibit satellite formation. The best known exceptions are Jupiter's gossamer rings¹ and Saturn's E ring, broad sheets of dust that extend outward until they fade from view at five to ten planetary radii. Source satellites continuously supply the dust, which is subsequently lost in collisions or by radial transport. Here we report that Saturn has an enormous ring associated with its outer moon Phoebe, extending from at least $128R_S$ to $207R_S$ (Saturn's radius R_S is 60,330 km). The ring's vertical thickness of $40R_S$ matches the range of vertical motion of Phoebe along its orbit. Dynamical considerations argue that these ring particles span the Saturnian system from the main rings to the edges of interplanetary space. The ring's normal optical depth of $\sim 2 \times 10^{-8}$ is comparable to that of Jupiter's faintest gossamer ring, although its particle number density is several hundred times smaller. Repeated impacts on Phoebe, from both interplanetary and circumplanetary particle populations, probably keep the ring populated with material. Ring particles smaller than centimetres in size slowly migrate inward^{2,3} and many of them ultimately strike the dark leading face of Iapetus.

In February 2009 we used the Spitzer Space Telescope's Multiband Imaging Photometer (MIPS) to scan regions near Phoebe's orbit to search for a broad debris ring. Mosaics produced from these mid-infrared images at $24\ \mu\text{m}$ and $70\ \mu\text{m}$ (MIPSON in Fig. 1) show a diffuse double-peaked band of light coincident with Saturn's ecliptic plane. The band is not produced by scattered light from Saturn (Fig. 2 and Supplementary Information) and has several characteristics consistent with grains that have spread into a ring after being liberated from Phoebe. The observed peak intensity of the ring at $24\ \mu\text{m}$ is $0.4\ \text{MJy sr}^{-1}$, about 1% of the zodiacal background. Particles with albedo 0.2 and emissivity 0.8 will have equilibrium temperature 85 K near Saturn, and a solid wall of them would emit with intensity $2 \times 10^6\ \text{MJy sr}^{-1}$ at $24\ \mu\text{m}$. The filling factor or line-of-sight optical depth of the ring is then simply the ratio of these two intensities: 2×10^{-7} . The optical depth would be an order of magnitude greater for bright icy grains at $70\ \text{K}$.

By far the largest of Saturn's distant satellites, Phoebe (mean radius 107 km; ref. 4) is probably the primary source of ejected debris in the outer Saturnian system. The moon follows an elliptical ($e = 0.16$) orbit around Saturn at an average distance of $a = 215R_S$. Phoebe's orbital plane is tilted by 5° from Saturn's, but because the moon travels in the direction opposite to that of the inner satellites its orbital inclination is $i = 175^\circ$. Particles launched from Phoebe share this tilt, and their orbital planes will precess with a characteristic time of thousands of years, producing a vertically extended torus with a full thickness of $2h = 2a(1 + e)\sin(i) = 41R_S$, a close match to the observed ring thickness (Fig. 3).

We have conducted numerical simulations of the evolution of different-sized dust grains ejected from Phoebe. Solar radiation pressure is the dominant perturbation force acting on particles up to tens of micrometres in size and it induces strong oscillations in orbital

eccentricities, efficiently spreading grains radially inward and outward (Fig. 4)⁵. Conversely, the $41R_S$ ring thickness is almost unaffected by radiation pressure for two reasons: (1) changes to orbital inclinations depend on the product $e\sin(i)$, which is always small; and (2) the greatest orbital tilts occur only when apocentres are in Saturn's orbital plane⁵. In Fig. 3b, the double peak is characteristic of rings supplied by inclined satellites¹, and the more distant ramps of the distribution are most probably due to the slight tilt of the ring relative to Spitzer's viewing direction.

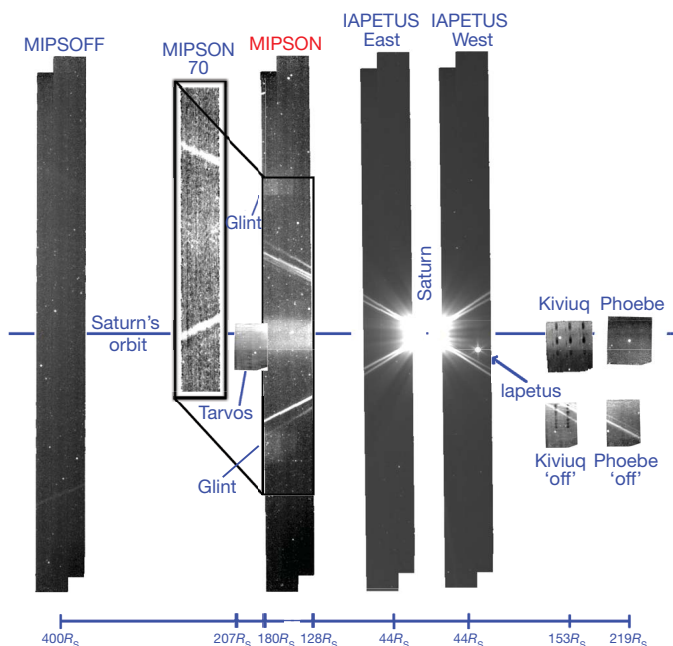


Figure 1 | Spitzer MIPS imaging in the vicinity of Saturn. The Phoebe ring appears as a bright emission feature centred in mosaics produced from scans at $24\ \mu\text{m}$ (MIPSON) and at a substantially lower signal-to-noise ratio at $70\ \mu\text{m}$ (MIPSON 70). At the closest point to Saturn, the edges of the MIPSON mosaic span radii from $128R_S$ to $180R_S$. The ring also appears serendipitously in the background of photometric observations of regions near outer satellites Kiviuq and Tarvos (centred at $153R_S$ and $190R_S$, respectively; Spitzer Program 03582, principal investigator T. Grav). A photometric observation of Phoebe (Spitzer Program 00071, principal investigator J. Houck) shows little diffuse emission, but the exposure was three times shorter and the ring brightness should be somewhat smaller at $220R_S$. No evidence for the ring appears in additional MIPS $24\ \mu\text{m}$ scans centred at $400R_S$ (MIPSOFF). The ring is overwhelmed by scattered light in $24\ \mu\text{m}$ scans (IAPETUS East and IAPETUS West) centred $44R_S$ from Saturn. Two off-target observations (Kiviuq 'off' and Phoebe 'off') illustrate typical background features in the absence of ring flux. The glints are artefacts that we describe in Fig. 2. See Supplementary Information for further observational details.

¹Department of Astronomy, University of Virginia, Charlottesville, Virginia, 22904-4325, USA, ²Department of Astronomy, University of Maryland, College Park, Maryland 20742-2421, USA.

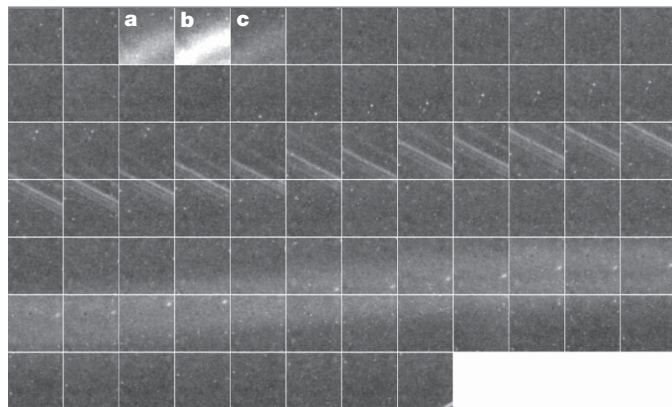


Figure 2 | A subset of the Spitzer MIPS 24 μm Basic Calibrated Data images, comprising a MIPSON scan. As a steadily scanning system, MIPS acquires a series of successive overlapping frames (in sequence from left to right) in which glints (Fig. 1) caused by scattered light appear and fade (a–c) over the course of a few frames and do not occupy a fixed location on the sky. These glints change geometry and brightness from frame to frame, as expected from a bright out-of-field source reflecting off different portions of the optical system. By contrast, the ring emission near Saturn's orbit plane (bottom panels) appears steady at constant flux and fixed on the sky in multiple MIPS frames.

These dynamical arguments allow us to calculate key ring parameters. Assuming a line of sight $500R_S$ through the ring at the point of observation, the cross-sectional area of all particles in a ring $300R_S$ in radius and $40R_S$ thick is $\sim 1.6 \times 10^7 \text{ km}^2$ (~ 500 times Phoebe's cross-section). These numbers depend only weakly on the ring's unknown particle size distribution; here we have assumed that $10 \mu\text{m}$ grains, the smallest that do not quickly reach Iapetus' orbit (Fig. 4), are predominant. A ring composed entirely of $10 \mu\text{m}$ grains would have number density $\sim 20 \text{ km}^{-3}$ and mass $\sim 3 \times 10^{11} \text{ kg}$, enough to fill a 1-km-diameter crater on Phoebe. In reality, collisions create a size distribution of ejecta, so the number density is likely to be appreciably lower and the ring mass correspondingly higher.

On long timescales, collisions and inward transport become important. Collision with Phoebe, the dominant loss mechanism for particles larger than several centimetres in size, takes on the order of 10^{10} years. Re-radiation of absorbed sunlight exerts an asymmetric force on dust grains, causing them to spiral in towards Saturn with a

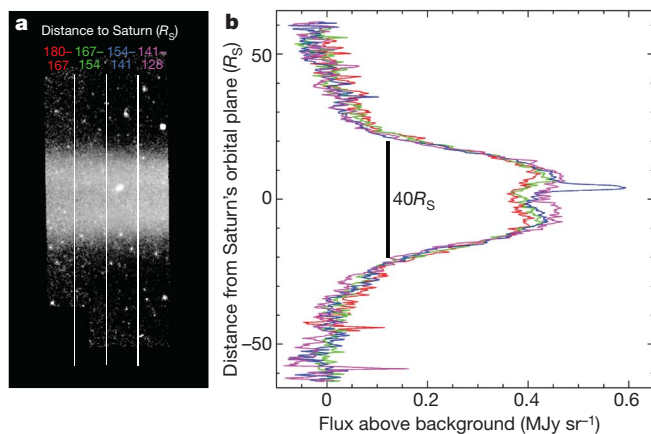


Figure 3 | Vertical profiles through the Phoebe dust structure. **a**, Colours denote radial ranges: magenta, $128\text{--}141R_S$; blue, $141\text{--}154R_S$; green, $154\text{--}167R_S$; and red, $167\text{--}180R_S$. **b**, The observed flux (MJy sr^{-1}) in four sets of forty averaged columns above and below Saturn's orbital plane produces a broad double peak 0.4 MJy sr^{-1} above the background, with a characteristic width of $40R_S$. The large spike near the peak of the blue profile is produced by the bright background galaxy readily visible in the MIPSON scan (**a**). The separation of the two peaks increases with radial distance from Saturn (**b**), as expected for a distribution of particle orbits with similar inclinations.

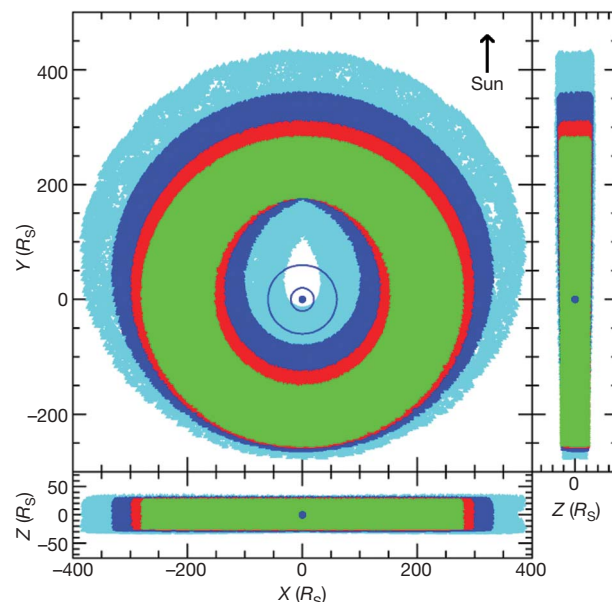


Figure 4 | The orbital distribution of dust grains launched from Phoebe followed for 2,000 years. Colours indicate particle radii in micrometres: 5 (cyan), 10 (blue), 20 (red), and 40 (green). The reference frame is centred on Saturn (blue dot) and rotates with the Sun's average angular motion. The X–Y plane is Saturn's orbital plane, and the Sun is located along the positive Y axis. Circles at $20R_S$ and $60R_S$ denote the orbits of Titan and Iapetus, respectively. Spitzer's view is approximately from the Sun and corresponds most closely to the X–Z plot. The numerical integrations assume a density of 1.6 g cm^{-3} , similar to that of Phoebe itself⁶, and include the gravity of Saturn, the Sun, Titan and Iapetus. To apply this figure to other assumed particle densities ρ , simply multiply all particle sizes by $1.6\rho^{-1}$. Solar radiation pressure, the dominant perturbation for small particles, forces the distribution of small grains to be azimuthally asymmetric and offset towards the Sun¹⁶. Conversely, grains $40 \mu\text{m}$ and larger form a symmetric torus around Saturn. Within 15 years (half a Saturn orbit), particles smaller than $3.5 \mu\text{m}$ will strike Saturn or its rings, while those smaller than $1.5 \mu\text{m}$ will be rapidly ejected from the Saturnian system.

characteristic timescale of $1.5 \times 10^5 r_g$ years where r_g is the particle radius in micrometres. This force brings all centimetre-sized and smaller material to Iapetus and Titan unless mutual particle collisions occur first. The rate of mutual collisions depends on the size distribution of the ring particles and optical depth; if the ring were comprised entirely of $10 \mu\text{m}$ grains, then the collisional timescale would be tens of millions of years, which is comparable to the inward drag timescale. Most material from $10 \mu\text{m}$ to centimetres in size ultimately hits Iapetus, with smaller percentages striking Hyperion and Titan³.

Many studies have suggested that dark material migrating inward from Phoebe and other outer Saturnian satellites coated the leading hemisphere of Iapetus^{2,3,6–11}. Recent observations by the Cassini spacecraft have revealed near-infrared spectral similarities between Phoebe and dark material on Iapetus and Hyperion^{12,13}, suggesting a common origin for dark material on the surfaces of all three. Dynamical studies predict that small irregular saturnian satellites have struck Phoebe several times over the age of the Solar System¹⁴, providing known historic sources of material, much of which remains in the ring today. The amount of dusty material currently in the ring is enough to cover the dark half of Iapetus to an average depth of $70 \mu\text{m}$. Interestingly, a ring composed of particles larger than our nominal $10 \mu\text{m}$ grains would be more massive, but would deliver that mass to Iapetus more slowly. As a result, the accumulation rate is relatively insensitive to the ring's unknown particle-size distribution. Assuming (1) that Iapetus intercepts all this material and (2) that the ring population is currently near its long-term average, the accumulation rate is about $40 \mu\text{m Myr}^{-1}$. Over the age of the Solar System, deposition at this rate would bury the

leading side of Iapetus to a depth of 20 cm. The population of distant satellites, however, was probably much higher in the past¹⁴, leading to more collisions, more debris and a cumulative thickness of material deposited on Iapetus that is probably measured in metres.

The closest analogues to the Phoebe ring are the two gossamer rings associated with Jupiter's inner satellites Thebe and Amalthea¹. These inner moons, while similar in size to Phoebe, are far more prolific sources of debris owing to more energetic collisions near Jupiter. Debris from Thebe re-impacts very rapidly (two thousand years) and so, ironically, the Phoebe and Thebe rings, which differ in scale by a factor of ~ 100 , actually have similar normal optical depths of $\sim 10^{-8}$. The new Saturnian structure is many hundred times thicker than the Jovian gossamer rings and its particle number density is correspondingly smaller. Its estimated mass in dust is many thousand times larger than that of either gossamer ring, especially when one accounts for the smaller particles in those rings¹⁵. In addition, Phoebe ring particles form a structure symmetric about Saturn's orbit plane, unlike all other known rings, which are equatorial. Finally, like Phoebe, these particles almost certainly follow retrograde orbits, moving clockwise when viewed from above, rather than counter-clockwise as do classical moons and rings. Although these exotic properties as well as its sheer size make the Phoebe ring unique among known planetary rings, similar structures should also adorn the other gas giant planets.

Received 21 August; accepted 18 September 2009.

Published online 7 October 2009.

1. Burns, J. A. *et al.* The formation of Jupiter's faint rings. *Science* **284**, 1146–1150 (1999).
2. Soter, S. Remarks on origin of Iapetus' photometric asymmetry. *IAU Colloq.* **28**, abstract (1974).
3. Burns, J. A., Hamilton, D. P., Mignard, F. & Soter, S. The contamination of Iapetus by Phoebe dust. *Astron. Soc. Pacif. Conf. Ser.* **104B**, 179–182 (1996).
4. Porco, C. C. *et al.* Cassini imaging science: initial results on Phoebe and Iapetus. *Science* **307**, 1237–1242 (2005).
5. Hamilton, D. P. Motion of dust in a planetary magnetosphere–orbit-averaged equations for oblateness, electromagnetic, and radiation forces with application to Saturn's E ring. *Icarus* **101**, 244–264 (1993).
6. Cruikshank, D. P. *et al.* The dark side of Iapetus. *Icarus* **53**, 90–104 (1983).
7. Buratti, B. J. & Mosher, J. A. The dark side of Iapetus: additional evidence for an exogenous origin. *Icarus* **115**, 219–227 (1995).
8. Vilas, F., Larsen, S. M., Stockstill, K. R. & Gaffey, M. J. Unraveling the zebra: clues to the Iapetus dark material composition. *Icarus* **124**, 262–267 (1996).
9. Jarvis, K. S., Vilas, F., Larsen, S. M. & Gaffey, M. J. Are Hyperion and Phoebe linked to Iapetus? *Icarus* **146**, 125–132 (2000).
10. Buratti, B. J., Hicks, M. D., Tryka, K. A., Sittig, M. S. & Newburn, R. L. High resolution 0.33–0.92 μm spectra of Iapetus, Hyperion, Phoebe, Rhea, Dione, and D-type asteroids: How are they related? *Icarus* **155**, 375–381 (2002).
11. Buratti, B. J., Hicks, M. D. & Davies, A. Spectrophotometry of the small satellites of Saturn and their relationship to Iapetus, Phoebe, and Hyperion. *Icarus* **175**, 490–495 (2005).
12. Cruikshank, D. P. *et al.* Hydrocarbons on Saturn's satellites Iapetus and Phoebe. *Icarus* **193**, 334–343 (2008).
13. Clark, R. N. *et al.* Compositional mapping of Saturn's moon Phoebe with imaging spectroscopy. *Nature* **435**, 66–69 (2005).
14. Nesvorný, D., Alvarro, J. L. A., Dones, L. & Levison, H. F. Orbital and collisional evolution of the irregular satellites. *Astron. J.* **126**, 398–429 (2003).
15. Hamilton, D. P., & Krüger, H. The sculpting of Jupiter's gossamer rings by its shadow. *Nature* **453**, 72–75 (2008).
16. Hamilton, D. P. The asymmetric time-variable rings of Mars. *Icarus* **119**, 153–172 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech.

Author Contributions All authors contributed substantially to this work. A.J.V. and M.F.S. did most of the observation planning, data reduction, and associated write-up. D.P.H. contributed most of the dynamical interpretations and write-up.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.J.V. (verbiscer@virginia.edu).

Information causality as a physical principle

Marcin Pawłowski¹, Tomasz Paterek², Dagomir Kaszlikowski², Valerio Scarani², Andreas Winter^{2,3}
& Marek Żukowski¹

Quantum physics has remarkable distinguishing characteristics. For example, it gives only probabilistic predictions (non-determinism) and does not allow copying of unknown states (no-cloning¹). Quantum correlations may be stronger than any classical ones², but information cannot be transmitted faster than light (no-signalling). However, these features do not uniquely define quantum physics. A broad class of theories exist that share such traits and allow even stronger (than quantum) correlations³. Here we introduce the principle of ‘information causality’ and show that it is respected by classical and quantum physics but violated by all no-signalling theories with stronger than (the strongest) quantum correlations. The principle relates to the amount of information that an observer (Bob) can gain about a data set belonging to another observer (Alice), the contents of which are completely unknown to him. Using all his local resources (which may be correlated with her resources) and allowing classical communication from her, the amount of information that Bob can recover is bounded by the information volume (m) of the communication. Namely, if Alice communicates m bits to Bob, the total information obtainable by Bob cannot be greater than m . For $m = 0$, information causality reduces to the standard no-signalling principle. However, no-signalling theories with maximally strong correlations would allow Bob access to all the data in any m -bit subset of the whole data set held by Alice. If only one bit is sent by Alice ($m = 1$), this is tantamount to Bob’s being able to access the value of any single bit of Alice’s data (but not all of them). Information causality may therefore help to distinguish physical theories from non-physical ones. We suggest that information causality—a generalization of the no-signalling condition—might be one of the foundational properties of nature.

Classical (as opposed to quantum) physics rests on the assumption that all physical quantities have well-defined values simultaneously. Relativity is based on clear-cut physical statements: the speed of light and the electric charge are the same for all observers. In contrast, the definition of quantum physics is still a description of its formalism: the theory in which systems are described by Hilbert spaces and dynamics is reversible. This situation is all the more unexpected because quantum physics is the most successful physical theory and quite a lot is known about it. Some of its counterintuitive features are almost popular knowledge: all scientists, and many laymen as well, know that quantum physics predicts only probabilities, that some physical quantities (such as position and momentum) cannot be simultaneously well defined and that the act of measurement generically modifies the state of the system. Entanglement and no-cloning are rapidly claiming their place in the list of well-known quantum features; in next place are the feats of quantum information such as the possibility of secure cryptography^{4,5} or the teleportation of unknown states⁶.

These features are so striking that one could hope that some of them provide the physical ground behind the formalism. Is quantum

physics, for instance, the most general theory that allows violations of Bell inequalities, while satisfying no-signalling? When this question was investigated³ the answer was found to be negative: impossibility of being represented in terms of local variables is a property shared by a broad class of no-signalling theories. Such theories predict intrinsic randomness, no-cloning^{7,8} and an information-disturbance trade-off⁹ and permit secure cryptography^{10–12}. As regards teleportation and entanglement swapping¹³, after a first negative attempt¹⁴, it seems that they can also be defined within the general no-signalling framework^{15,16}. In summary, most of the features that have been highlighted as ‘typically quantum’ are shared by all possible no-signalling theories. Only a few discrepancies have been noticed: some no-signalling theories would lead to an implausible simplification of distributed computational tasks^{17–20} and would have very limited dynamics²¹. This highlights the importance of the no-signalling principle but leaves us still uncertain about the specificity of quantum theory.

Here we define and study a previously unnoticed feature, which we call ‘information causality’. Information causality generalizes no-signalling and is respected by both classical and quantum physics. However, as we shall show, it is violated by all no-signalling theories that are endowed with correlations that are stronger than the strongest quantum correlations. It can therefore be used as a principle to distinguish physical theories from non-physical ones and is a good candidate for one of the foundational assumptions that are at the very root of quantum theory.

Formulated as a principle, information causality states: “the information gain that Bob can reach about a previously unknown to him data set of Alice, by using all his local resources and m classical bits communicated by Alice, is at most m bits”. The standard no-signalling condition is just information causality for $m = 0$. The principle assumes classical communication: if quantum bits were allowed to be transmitted, the information gain could be higher, as demonstrated in the quantum super-dense coding protocol²². The efficiency of this protocol is based on the use of quantum entanglement, and information causality holds true even if the quantum bits are transmitted provided that they are disentangled from the systems of the receiver. This follows from the Holevo bound, which limits information gain after transmission of m such qubits to m classical bits.

We show that in a world in which certain tasks are ‘too simple’ (compare with refs 17, 18) and there exists implausible accessibility of remote data, information causality is violated. Consider a generic situation in which Alice has a database of N bits described by a string \vec{a} . She would like to grant Bob access to as big a portion of the database as possible within a fixed amount of classical communication. If there were no pre-established correlations between them, communication of m bits would open access to at most m bits of the database. With previously shared correlations they could expect to do better (however, as we show here, in the real world they would be mistaken). For concreteness, consider a generic task illustrated in Fig. 1. It is a

¹Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland. ²Centre for Quantum Technologies and Department of Physics, National University of Singapore, 3 Science Drive 2, 117543 Singapore, Singapore. ³Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK.

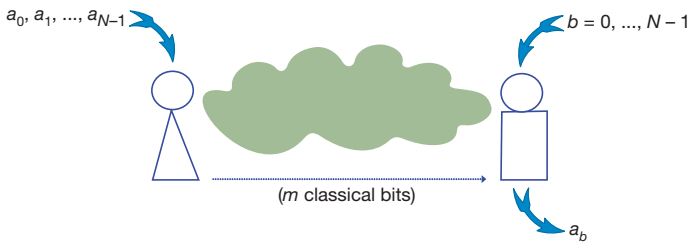


Figure 1 | The task. Alice receives N random and independent bits $\vec{a} = (a_0, \dots, a_{N-1})$. In a separate location, Bob receives a random variable $b \in \{0, 1, \dots, N-1\}$. Alice sends m classical bits to Bob, with the help of which Bob is asked to guess the value of the b th bit in Alice's list, a_b . Alice and Bob can share any no-signalling resources. Information causality limits the efficiency of solutions to this task. It bounds the mutual information between Alice's data and all that Bob has at hand after receiving the message.

distributed version of random access coding^{23,24}, oblivious transfer^{14,25} and related communication complexity problems²⁶. Alice receives a string of N random and independent bits, $\vec{a} = (a_0, a_1, \dots, a_{N-1})$. Bob receives a random value of $b = 0, \dots, N-1$ and is asked to give a value of the b th bit of Alice after receiving from her a message of m classical bits. The restrictions are only on the communication that can take place after the inputs have been provided. The resources that Alice and Bob may have shared in advance are assumed to be no-signalling because allowing signalling resources would open other communication channels. In a classical world, these additional resources would be correlated lists of bits; in a quantum world, Alice and Bob may share an arbitrary quantum state. However, the task itself is open to accommodate any hypothetical resource producing no-signalling correlations, even those that go beyond the possibilities of quantum physics. We shall call these imaginary resources no-signalling boxes, or NS-boxes for short. The impact of stronger-than-quantum correlations on the efficiency of random access coding has been studied recently from a different angle²⁴.

There exists a protocol that allows Bob to give the correct value of at least m bits. If Alice sends him an m -bit message $\vec{x} = (a_0, \dots, a_{m-1})$ Bob will guess a_b perfectly whenever $b \in \{0, \dots, m-1\}$. The price to pay is that he is bound to make a completely random guess for $b \in \{m, \dots, N-1\}$. Because the previously shared correlations contain no information about \vec{a} , for every strategy there will be a trade-off between the probabilities for guessing different bits of \vec{a} . Let us denote Bob's output by β . The efficiency of Alice's and Bob's strategy can be quantified by

$$I \equiv \sum_{K=0}^{N-1} I(a_K : \beta | b=K) \quad (1)$$

where $I(a_K : \beta | b=K)$ is the Shannon mutual information between a_K and β , computed under the condition that Bob has received $b=K$. One can also show that

$$I \geq N - \sum_{K=0}^{N-1} h(P_K) \quad (2)$$

where $h(x) = -x \log_2 x - (1-x) \log_2 (1-x)$ is the binary entropy of x , and P_K is the probability that $a_K = \beta$, again for the case of $b=K$. To obtain the inequality, the a_K have been assumed to be unbiased and independently distributed (details are given in Supplementary Information).

Ideally, we wish to define that information causality holds if, after transfer of the m -bit message, the mutual information between Alice's data \vec{a} and everything that Bob has—that is, the message \vec{x} and his part B of the previously shared correlation—is bounded by m . Intuitively appealing though such a definition is, it has the severe issue that it is not theory-independent. Specifically, a mutual information expression ' $I(\vec{a} : \vec{x}, B)$ ' has to be defined for a state involving objects from the underlying theory (the possibilities include classical correlation, a

shared quantum state and NS-boxes). It is far from clear whether mutual information can be defined consistently for all non-local correlations, nor whether such a definition would be unique.

Instead, we shall show that if a mutual information can be defined that obeys certain elementary properties, then (a) information causality holds and (b) $I(\vec{a} : \vec{x}, B) \geq I$. Thus we obtain the following necessary condition for information causality:

$$I \leq m \quad (3)$$

We stress that the parameter I is independent of any underlying physical theory: I does not involve any details of a particular physical model but is fully determined by Alice's and Bob's input bits and Bob's output. In this sense it resembles Bell's parameter², which also involves only random variables and can be used to test different physical theories.

For a system composed of parts A , B and C , prepared in a state allowed by the theory, we need to assign symmetric and non-negative mutual informations $I(A : B)$, etc. The elementary properties mentioned above are the following. First, consistency: if the subsystems A and B are both classical, then $I(A : B)$ should coincide with Shannon's mutual information. Second, data-processing inequality: acting on one of the parts locally by any state transformation allowed in the theory cannot increase the mutual information. That is, if $B \rightarrow B'$ is a permissible map between systems, then $I(A : B) \geq I(A : B')$. This says that any local manipulation of data can only decay information. Third, a chain rule: there exists a conditional mutual information $I(A : B | C)$ such that the following identity is satisfied for all states and triples of parts: $I(A : B, C) = I(A : C) + I(A : B | C)$. This implies an identity between ordinary mutual informations:

$$I(A : B, C) - I(A : B) = I(A : B | C) = I(A, C : B) - I(B : C)$$

Information causality holds in both classical and quantum physics; we may focus on the latter because the former is a special case of it. This is because one can define quantum mutual information in a formal extension of Shannon's quantity, using von Neumann entropy²⁷, and all three of the above properties are fulfilled²⁸. Details are given in Supplementary Information, but in brief one argues as follows.

To show (a), denote by B Bob's quantum system holding the shared quantum state ρ_{AB} , Alice's data $\vec{a} = (a_0, \dots, a_{N-1})$, and the m -bit message \vec{x} ; our objective is to prove that $I(\vec{a} : \vec{x}, B) \leq m$. First, the chain rule for mutual information yields $I(\vec{a} : \vec{x}, B) = I(\vec{a} : B) + I(\vec{a} : \vec{x} | B)$. Second, $I(\vec{a} : B) = 0$ because without the message Alice's data and Bob's quantum state are independent (expressing the no-signalling condition). Third, we use the chain rule again to express the conditional mutual information as $I(\vec{a} : \vec{x} | B) = I(\vec{x} : \vec{a}, B) - I(\vec{x} : B) \leq I(\vec{x} : \vec{a}, B)$. Finally, the latter can be upper-bounded by $I(\vec{x} : \vec{x}) \leq m$, invoking data processing. Similarly, (b) is obtained by repeated application of the chain rule, data-processing inequality and non-negativity of mutual information (details are given in Supplementary Information).

To study how other no-signalling theories can violate information causality, we focus on the necessary condition in equation (3). First consider the simplest example of two-bit input by Alice, (a_0, a_1) ; it is described in Fig. 2. The probability that Bob correctly gives the value of the bit a_0 is

$$P_I = \frac{1}{2} [P(A \oplus B = 0 | 0, 0) + P(A \oplus B = 0 | 1, 0)] \quad (4)$$

and the analogous probability for the bit a_1 reads

$$P_{II} = \frac{1}{2} [P(A \oplus B = 0 | 0, 1) + P(A \oplus B = 1 | 1, 1)] \quad (5)$$

where the symbol \oplus denotes summation modulo 2.

One can recognize that these probabilities are intimately linked with the Clauser–Horne–Shimony–Holt parameter²⁹ S , which can be used to quantify the strength of correlations. Indeed,

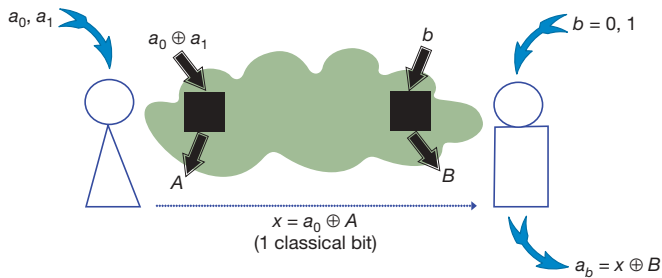


Figure 2 | Van Dam's protocol¹⁷. This is the simplest case in which information causality can be violated (see also ref. 25). Alice receives two bits (a_0, a_1) and is allowed to send only one bit to Bob. A convenient way of thinking about no-signalling resources is to consider paired black boxes shared between Alice and Bob (NS-boxes). The correlations between inputs $a, b = 0, 1$ and outputs $A, B = 0, 1$ of the boxes are described by probabilities $P(A \oplus B = ab|a, b)$. The no-signalling is satisfied because of uniformly random local outputs. With suitable NS-boxes Alice and Bob violate information causality. She uses $a = a_0 \oplus a_1$ as an input to the shared NS-box and obtains the outcome A , which is used to compute her message bit $x = a_0 \oplus A$ for Bob. Bob, on his side, inputs $b = 0$ if he wants to learn a_0 , and $b = 1$ if he wants to learn a_1 ; he gets the outcome B . On receiving x from Alice, Bob computes his guess $\beta = x \oplus B = a_0 \oplus A \oplus B$. The probability that Bob correctly gives the value of the bit a_0 is $P_I = \frac{1}{2}[P(A \oplus B = 0|0, 0) + P(A \oplus B = 0|1, 0)]$, and the analogous probability for the bit a_1 reads $P_{II} = \frac{1}{2}[P(A \oplus B = 0|0, 1) + P(A \oplus B = 1|1, 1)]$, which follow by inspection of the different cases.

$$S = \sum_{a=0}^1 \sum_{b=0}^1 P(A \oplus B = ab|a, b) = 2(P_I + P_{II}) \quad (6)$$

The classical correlations are bounded by $S \leq S_C = 3$ (the equivalent form of Bell's inequality^{2,29}). Quantum correlations exceed this limit up to $S \leq S_Q = 2 + \sqrt{2}$ (the so-called Tsirelson bound³⁰). The maximal algebraic value of $S_{NS} = 4$ is reached by the Popescu–Rohrlich (PR) box³, which is an extremal no-signalling resource. PR-boxes maximally violate information causality because they predict $P_I = P_{II} = 1$; that is, $I = 2$ for $m = 1$, so here occurs an extreme violation of information causality. Bob can learn either bit perfectly. $I = 2$ measures the sum total of the information accessible to Bob. However, he cannot learn both of Alice's bits—the latter would imply signalling.

The protocol works just as well for any Boolean function of the inputs, $f(\vec{a}, b)$. It is sufficient that Alice inserts to her PR-box the sum $f(\vec{a}, 0) \oplus f(\vec{a}, 1)$. If information causality is maximally violated, Bob can learn the value of $f(\vec{a}, b)$ for any one of his inputs, irrespective of Alice's input data. Even more surprisingly, this is also true if he does not know the function to be computed.

We shall now demonstrate that information causality is violated as soon as the quantum Tsirelson limit for the CHSH inequality is exceeded. This result of ours can be also seen as an information-theoretic proof of the Tsirelson bound, independently of the formalism of Hilbert spaces, relying instead only on the existence of a consistent information calculus for certain correlations.

First we note that, using a suitable local randomization procedure that does not change the value of the parameter S , any NS-box can be brought to a simple form⁷: the local outcomes are uniformly random and the correlations are given by

$$P(A \oplus B = ab|a, b) = \frac{1}{2}(1 + E) \quad (7)$$

with $0 \leq E \leq 1$. The case $E = 1$ corresponds to the PR-box; $E = 0$ describes uncorrelated random bits. The classical bound $S \leq S_C$ is violated as soon as $E > \frac{1}{2}$; the Tsirelson bound of quantum physics becomes $E \leq E_Q = \frac{1}{\sqrt{2}}$, attained by performing suitable measurements on the singlet state of two two-level systems^{2,30}.

The bound that information causality imposes on correlations can be identified by using a pyramid of NS-boxes and nesting the simple

protocol described above. Now Alice receives $N = 2^n$ bits, and correspondingly Bob receives n input bits b_m , which describe the index of the bit he is interested in, $b = \sum_{k=0}^{n-1} b_k 2^k$. Alice is allowed to send a single bit, $m = 1$. An example of this protocol for $n = 2$ is presented in Fig. 3. Generally, Alice and Bob use a pyramid of $N - 1$ pairs of boxes placed on n levels. Looking at the binary decomposition of b , Bob aims $(n - r)$ times at the left bit and r times at the right, where $r = b_0 + \dots + b_{n-1}$. His final guess is the sum of $\beta = x \oplus B_0 \oplus \dots \oplus B_{n-1}$. Bob's final guess is therefore correct whenever he has made an even number of errors in the intermediate steps. This leads to the equation

$$P_K = \frac{1}{2}[1 + E^n] \quad (8)$$

for the probability of his correct final guess (the details of this calculation are given in Supplementary Information).

Inserting this expression into equation (1), one finds that the information causality condition $I \leq 1$ is violated as soon as $2E^2 > 1$ and n is large enough; that is, $E > E_Q$. Because all NS-boxes can be brought to the form in equation (7) without changing the value of S , we conclude indeed that every NS-box with stronger than quantum correlations violates the information causality condition. In Supplementary Information the more general result is proved, that for any $\frac{1}{2}(E_1^2 + E_2^2) > E_Q^2$ where $E_j = 2P_j - 1$ (see equations (4) and (5)), information causality is violated, and conversely that if it is fulfilled there exists a quantum correlation with these probabilities.

Here we have identified the principle of information causality, which precisely distinguishes physically realized correlations from non-physical ones (in the sense that quantum mechanics cannot reach them). It is phrased in operational terms and in a theory-independent way; we therefore suggest that it is at the same foundational level as the no-signalling condition itself, of which it is a generalization.

The new principle is respected by all correlations accessible with quantum physics and excludes all no-signalling correlations, which violate the quantum Tsirelson bound. Among the correlations that do not violate that bound it is not known whether information

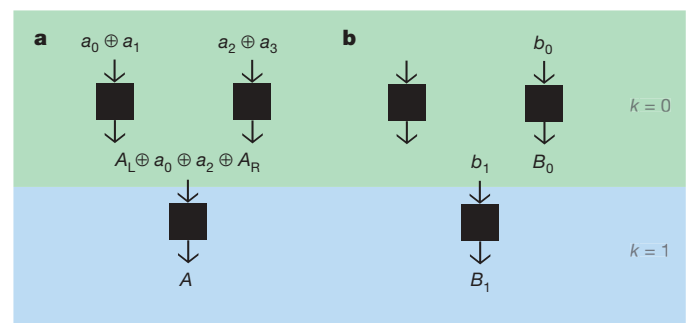


Figure 3 | Information causality identifies the strongest quantum correlations. Alice receives $N = 2^n$ input bits and correspondingly Bob receives n input bits b_m , which describe the index of the bit he is interested in, $b = \sum_{k=0}^{n-1} b_k 2^k$. Alice is allowed to send a single bit, $m = 1$. **a**, For $n = 2$, to encode information about her data Alice uses a pyramid of NS-boxes. Note that Fig. 2 shows how Bob can correctly guess the first or second bit Alice has using a single pair of the boxes (the case of $n = 1$). If Alice has more bits, then they recursively use this protocol in the following way. For example, for four input bits of Alice, two pairs of NS-boxes on the level $k = 0$ allow Bob to make the guess of a value of any one of Alice's bits as soon as he knows either $a_0 \oplus A_L$ or $a_2 \oplus A_R$, where A_L and A_R are the output of her left and right boxes, respectively, on the level $k = 0$, which are the one-bit messages of the protocol in Fig. 2. These can be encoded using the third box, on the level $k = 1$, by inserting their sum to Alice's box and sending $x = a_0 \oplus A_L \oplus A$ to Bob (A is the output of her box on the level $k = 1$). Depending on the bit he is interested in, he now reads a suitable message using the box on the level $k = 1$ and uses one of the boxes on the level $k = 0$. **b**, An example of a situation in which Bob aims at the value of a_2 or a_3 . Bob's final answer is $x \oplus B_0 \oplus B_1$, where B_k is the output of his box on the k th level.

causality singles out exactly those allowed by quantum physics. If it does, the new principle would acquire even stronger status.

Received 8 May; accepted 13 August 2009.

- Wooters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
- Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200 (1964).
- Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379–383 (1991).
- Bennett, C. H. & Brassard, G. in *Proceedings IEEE Int. Conf. on Computers, Systems and Signal Processing, Bangalore, India* 175–179 (IEEE, 1984).
- Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Masanes, L., Acín, A. & Gisin, N. General properties of nonsignaling theories. *Phys. Rev. A* **73**, 012112 (2006).
- Barnum, H., Barrett, J., Leifer, M. & Wilce, A. Generalized no-broadcasting theorem. *Phys. Rev. Lett.* **99**, 240501 (2007).
- Scarani, V. *et al.* Secrecy extraction from no-signaling correlations. *Phys. Rev. A* **74**, 042339 (2006).
- Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Phys. Rev. Lett.* **95**, 010503 (2005).
- Acín, A., Gisin, N. & Masanes, L. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.* **97**, 120405 (2006).
- Masanes, L. Universally-composable privacy amplification from causality constraints. *Phys. Rev. Lett.* **102**, 140501 (2009).
- Żukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. 'Event-ready-detectors' Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993).
- Short, A. J., Popescu, S. & Gisin, N. Entanglement swapping for generalized nonlocal correlations. *Phys. Rev. A* **73**, 012101 (2006).
- Barnum, H., Barrett, J., Leifer, M. & Wilce, A. Teleportation in general probabilistic theories. Preprint at <http://arxiv.org/abs/0805.3553v1> (2008).
- Skrzypczyk, P., Brunner, N. & Popescu, S. Emergence of quantum correlations from nonlocality swapping. *Phys. Rev. Lett.* **102**, 110402 (2009).
- Van Dam, W. *Nonlocality and Communication Complexity*. PhD thesis, Univ. Oxford (2000); Implausible consequences of superstrong nonlocality. Preprint at <http://arxiv.org/abs/quant-ph/0501159v1> (2005).
- Brassard, G. *et al.* Limit on nonlocality in any world in which communication complexity is not trivial. *Phys. Rev. Lett.* **96**, 250401 (2006).
- Linden, N., Popescu, S., Short, A. J. & Winter, A. Quantum nonlocality and beyond: limits from nonlocal computation. *Phys. Rev. Lett.* **99**, 180502 (2007).
- Brunner, N. & Skrzypczyk, P. Non-locality distillation and post-quantum theories with trivial communication complexity. *Phys. Rev. Lett.* **102**, 160403 (2009).
- Barrett, J. Information processing in generalized probabilistic theories. *Phys. Rev. A* **75**, 032304 (2007).
- Bennett, C. H. & Wiesner, S. J. Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states. *Phys. Rev. Lett.* **69**, 2881–2884 (1992).
- Ambainis, A., Nayak, A., Ta-Shma, A. & Vazirani, U. Quantum dense coding and quantum finite automata. *J. ACM* **49**, 496–511 (2002).
- Ver Steeg, G. & Wehner, S. Relaxed uncertainty relations and information processing. Preprint at <http://arxiv.org/abs/0811.3771v2> (2009).
- Wolf, S. & Wullschleger, J. Oblivious transfer and quantum non-locality. Preprint at <http://arxiv.org/abs/quant-ph/0502030v1> (2005).
- Brassard, G. Quantum communication complexity. *Found. Phys.* **33**, 1593–1616 (2003).
- Cerf, N. J. & Adami, C. Negative entropy and information in quantum mechanics. *Phys. Rev. Lett.* **79**, 5194–5197 (1997).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- Cirel'son, B. S. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4**, 93–100 (1980).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Christandl, V. Vedral and S. Wehner for stimulating discussions. This work was supported by the National Research Foundation and the Ministry of Education in Singapore, and by the European Commission through the Integrated Project Qubit Applications. A.W. acknowledges support by the UK Engineering and Physical Sciences Research Council through the Quantum Information Processing Interdisciplinary Research Collaboration and an Advanced Fellowship, by a Royal Society Wolfson Merit Award, and by a Philip Leverhulme Prize.

Author Contributions All authors contributed to the initial conception of the ideas, to the working out of details, and to the writing and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.P. (dokmpa@univ.gda.pl).

Imaging chromophores with undetectable fluorescence by stimulated emission microscopy

Wei Min^{1*}, Sijia Lu^{1*}, Shasha Chong¹, Rahul Roy¹, Gary R. Holtom¹ & X. Sunney Xie¹

Fluorescence, that is, spontaneous emission, is generally more sensitive than absorption measurement, and is widely used in optical imaging^{1,2}. However, many chromophores, such as haemoglobin and cytochromes, absorb but have undetectable fluorescence because the spontaneous emission is dominated by their fast non-radiative decay³. Yet the detection of their absorption is difficult under a microscope. Here we use stimulated emission, which competes effectively with the nonradiative decay, to make the chromophores detectable, and report a new contrast mechanism for optical microscopy. In a pump-probe experiment, on photoexcitation by a pump pulse, the sample is stimulated down to the ground state by a time-delayed probe pulse, the intensity of which is concurrently increased. We extract the miniscule intensity increase with shot-noise-limited sensitivity by using a lock-in amplifier and intensity modulation of the pump beam at a high megahertz frequency. The signal is generated only at the laser foci owing to the nonlinear dependence on the input intensities, providing intrinsic three-dimensional optical sectioning capability. In contrast, conventional one-beam absorption measurement exhibits low sensitivity, lack of three-dimensional sectioning capability, and complication by linear scattering of heterogeneous samples. We demonstrate a variety of applications of stimulated emission microscopy, such as visualizing chromoproteins, non-fluorescent variants of the green fluorescent protein, monitoring *lacZ* gene expression with a chromogenic reporter, mapping transdermal drug distributions without histological sectioning, and label-free microvascular imaging based on endogenous contrast of haemoglobin. For all these applications, sensitivity is orders of magnitude higher than for spontaneous emission or absorption contrast, permitting non-fluorescent reporters for molecular imaging.

The phenomenon of stimulated emission was first described by Einstein in 1917 (ref. 4). An atom or molecule in its excited state can be stimulated down to the ground state by an incident light field, resulting in the creation of a new coherent photon identical to those in the original incident field. This process only occurs when the frequency of the incident field matches the transition energy. Stimulated emission was later used as a fundamental principle for light amplification in the laser⁵. The depopulation aspect of stimulated emission has been used for population dumping from excited states⁶, super-resolution fluorescence microscopy⁷, and fluorescence lifetime imaging⁸. Here we use the light-amplification aspect of stimulated emission as a contrast mechanism for highly sensitive imaging of chromophores that have undetectable fluorescence.

Such chromophores have very short-lived excited states with much faster non-radiative decay rates than their spontaneous emission rates. As a result, their feeble fluorescence is overwhelmed by backgrounds, such as stray light, solvent Raman scattering, and detector dark counts. Our solution to this problem is to conduct a dual-beam

experiment to interrogate the short-lived excited state by stimulated emission, which can compete with the non-radiative decay under a strong enough stimulating field (Fig. 1a). The resulting ‘amplification’ of the stimulation beam can then be detected in the presence of the background signals.

Considering the optical excitation at frequency ω_{01} (Fig. 1a), the absorption cross-section $\sigma_{\text{abs}[0 \rightarrow 1]}$ is about 10^{-16} cm^2 for a single chromophore at room temperature^{2,9}. As shown in Fig. 1b, under a tightly focused laser beam with a beam waist area of S ($\sim 10^{-9} \text{ cm}^2$), the integrated intensity attenuation of the excitation beam $\Delta I_E/I_E$ is proportional to the ratio between $\sigma_{\text{abs}[0 \rightarrow 1]}$ and S :

$$\Delta I_E/I_E = -N_0 \sigma_{\text{abs}[0 \rightarrow 1]}/S \quad (1)$$

where N_0 is the number of molecules in the ground state. For a single chromophore, $\Delta I_E/I_E$ is of the order of 10^{-7} . Such small attenuation cannot be detected by conventional absorption microscopy. We note that single-molecule absorption was previously achieved in cryogenic temperatures using a frequency-modulation method¹⁰, which is, however, difficult to implement at room temperature because of the broad molecular absorption linewidth. Moreover, absorption measurement is often complicated by scattering when imaging biological specimens. Instead of detecting direct absorption, here we detect the stimulated emission followed by the excitation of the molecule.

According to Einstein⁴, the molecular cross-section σ_{stim} for stimulated emission is comparable to σ_{abs} , because of microscopic reversibility. Unlike the absorption that results in attenuation, the stimulation beam will experience an intensity gain after interacting with the molecules:

$$\Delta I_S/I_S = +N_2 \sigma_{\text{stim}[2 \rightarrow 3]}/S \quad (2)$$

where N_2 is the number of excited molecules transiently probed by the stimulation pulses (Fig. 1b). For a single chromophore, $\Delta I_S/I_S$ is also $\sim 10^{-7}$. Without special techniques, such a small signal would again be buried in the noise ($\sim 1\%$) of the stimulation beam.

To overcome this noise problem in detecting stimulated emission, we implemented a high-frequency ($>1 \text{ MHz}$) phase-sensitive detection technique. In so doing, the laser intensity fluctuation, which occurs primarily at low frequency (kilohertz to direct current), can be circumvented, as has been previously applied in other spectroscopic¹¹ and recently stimulated Raman scattering microscopy¹² and two-photon absorption microscopy¹³. In the scheme shown in Fig. 1b and c, the intensity of the excitation beam is modulated at 5 MHz, and this creates a modulation of the stimulated emission signal at the same frequency, because only when the excitation beam is present can the gain of the stimulation beam occur. Such an induced modulation signal can then be sensitively extracted by a lock-in amplifier referenced

¹Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

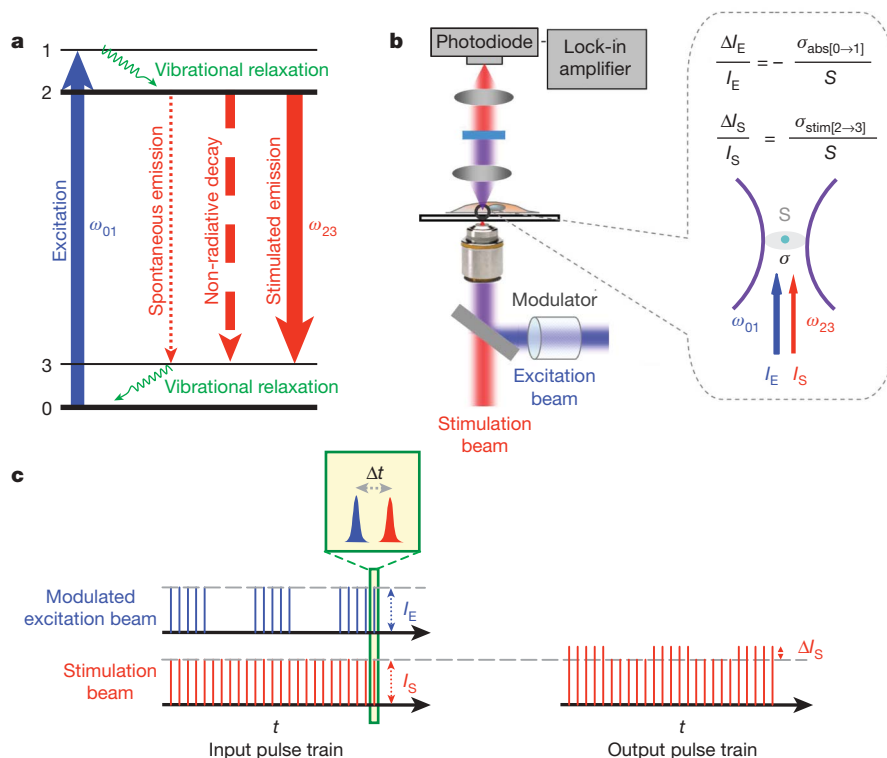


Figure 1 | Principles of stimulated emission microscopy. **a**, Energy diagram of spontaneous emission, non-radiative decay and stimulated emission processes for a typical four-level energy system. Non-radiative decay dominates spontaneous emission for chromophores with undetectable fluorescence. However, when the stimulation field is designed to have the correct energy and timing, the stimulated emission can compete with non-radiative decay and become the dominating decay pathway. **b**, Stimulated emission microscopy. The incident excitation and delayed stimulation pulse trains are spatially overlapped and focused onto the common focal spot in the sample. A modulator switches the intensity of the excitation beam on and

to this high frequency. In this way, the stimulated emission signal can be detected against the laser noise with orders-of-magnitude higher amplitude. In contrast, it is difficult to implement such a modulation scheme for high-sensitivity one-beam absorption measurement¹¹.

Specifically, we use a ~ 200 -femtosecond (fs) (full-width at half-maximum, FWHM) pulse train for excitation, and another ~ 200 -fs pulse train for stimulation. The time delay between these two pulse trains is chosen to be ~ 300 fs (Fig. 1c), which is shorter than the excited-state lifetime (sub-picosecond, ps) of the chromophores. This delay also eliminates contributions from other instantaneous optical processes, such as two-photon absorption¹³, cross-phase modulation and stimulated Raman scattering¹².

We conduct the experiment under a non-saturating condition of the four-level system (Fig. 1a). Under this condition, N_2 in equation (2) originates from a linear excitation: $N_2 \propto N_0 I_E \sigma_{\text{abs}[0 \rightarrow 1]} / S$. This relation, together with equation (2), indicates that the final signal ΔI_S is linearly dependent on both I_E and I_S :

$$\Delta I_S \propto N_0 I_E I_S (\sigma_{\text{abs}[0 \rightarrow 1]} / S) / (\sigma_{\text{stim}[2 \rightarrow 3]} / S) \quad (3)$$

The overall quadratic power dependence, as experimentally demonstrated (Fig. 2a), would allow three-dimensional optical sectioning, as in many other multi-photon techniques^{14,15}. Such dependence also confirms the non-saturation condition within the intensity range used. Moreover, it offers, in principle, a spatial resolution twice as high (in spatial frequency) as in conventional fluorescence microscopy.

Figure 2b shows the stimulated emission signal as a function of the time delay between the excitation and stimulation pulses. In the pump-probe experiment, the signal could arise from ground-state

off at 5 MHz. The spectrally filtered stimulation beam is detected by a large-area photodiode, and demodulated by a lock-in amplifier to create the image contrast while raster scanning the collinear exciting and stimulating beams. The relative energy gain or loss of the stimulation and excitation beam, respectively, for a single chromophore at the laser focus (area of S), is given by the equations (inset). **c**, Principles of high-frequency modulation. As the intensity of the excitation beam is modulated on and off over time, the corresponding stimulated emission signal is also modulated at the same frequency. The stimulation pulses are made to lag behind the excitation pulses by an appropriate Δt .

recovery, excited-state absorption or stimulated emission. Ground-state recovery is avoided here by the large Stokes shift of stimulation beam from the molecular absorption band. Excited-state absorption is negligible compared to stimulated emission, on the basis of the observed intensity increase of the stimulation beam, consistent with a study of fluorescence quenching in stimulated emission depletion microscopy¹⁶. The initial rise in Fig. 2b is due to vibrational relaxation from level 1 to level 2 (Fig. 1a), while the subsequent decay indicates the short excited-state lifetime (~ 0.6 ps) of level 2, which underlies the undetectable fluorescence. Such a short lifetime also reduces the probability of going into the triplet state, effectively protecting the molecule from photo-bleaching. The stimulated emission spectrum (Fig. 2c), recorded by tuning the wavelength of the stimulated beam, is also in agreement with the reported emission spectrum of crystal violet in glycerol¹⁷, in which the high viscosity increases the fluorescence quantum yield.

Figure 2d shows that the stimulated emission signal depends linearly on analyte concentration, as predicted by equation (3). This allows straightforward quantitative analysis. The limit of detection is ~ 60 nM for crystal violet with 1 s integration time. This sensitivity effectively corresponds to a few (< 5) molecules in focus, which has surpassed the detection limit recently reported for stimulated Raman scattering microscopy¹² by orders of magnitude because of the strong electronic transition. The ultimate sensitivity would be achieved under the condition of saturated excitation but near-saturated stimulation, and is fundamentally determined by the shot noise of the stimulation beam ($\Delta I_S / I_S \approx 10^{-7.5}$ for ~ 1 mW, averaged over 1 s).

As the first biological application, we imaged distributions of chromoproteins in live *Escherichia coli* cells. Genetically encodable

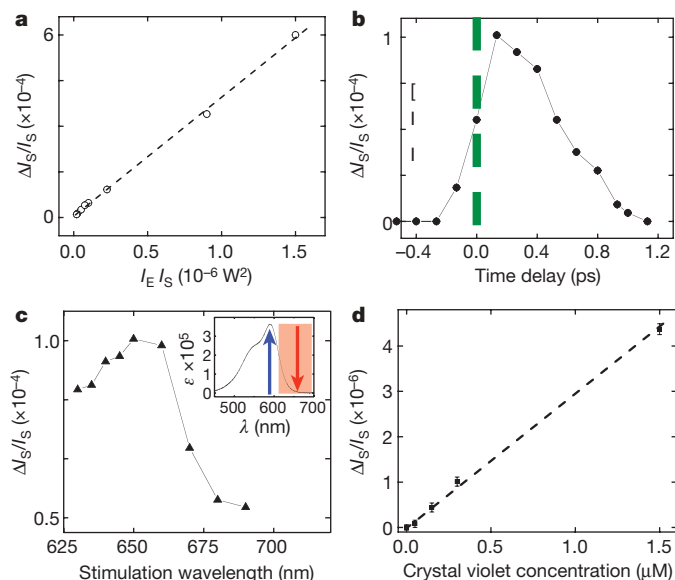


Figure 2 | Characterizations of stimulated emission microscopy.

a, Stimulated emission signal from 30 μM crystal violet/water solution is proportional to the product of excitation beam power, I_E , and stimulation beam power, I_S . **b**, Stimulated emission signal as a function of time delay between the excitation (590 nm) and the stimulation (660 nm) pulses. The decay of the signal (~ 0.6 ps) reflects the short excited-state lifetime. The FWHM of the pulses is ~ 200 fs. **c**, The measured stimulated emission spectrum is in agreement with the reported fluorescence spectrum of crystal violet in glycerol solution¹⁶. The 590 nm excitation is fixed while the stimulation wavelength is scanned by tuning the other optical parametric oscillator wavelength. A time delay of 0.3 ps is used. Inset, the absorption spectrum of crystal violet in water. ϵ , Molar extinction coefficient. **d**, The stimulated emission signal scales linearly with crystal violet concentration in aqueous solution. Continuous flow of the sample is used to replenish molecules. Wavelengths are 590 nm and 660 nm for excitation and stimulation beam, respectively, and the time delay is 0.3 ps. Error bars show 1 s.d. of the signals from a 30 s recording. The detection limit was determined to be 60 nM with a signal-to-noise ratio of 1:1 (Supplementary Fig. 3). The excitation and stimulation beam intensities at the focus are 0.2 mW ($\sim 0.14 \text{ MW cm}^{-2}$) and 1 mW ($\sim 0.70 \text{ MW cm}^{-2}$), respectively. For 1 s time constant at the lock-in amplifier, a relative signal of 10^{-7} for $\Delta I_S/I_S$ can be detected.

chromoproteins, such as gtCP¹⁸ and cjBlue¹⁹, are naturally existing variants of green fluorescent proteins²⁰; they absorb light but do not fluoresce. When the gene encoding gtCP is expressed in *E. coli* cells, tetrameric gtCP can be clearly seen residing homogeneously inside the cytoplasm by stimulated emission microscopy (Fig. 3a and b). Similarly, the distribution of another chromoprotein, cjBlue, can be imaged (Fig. 3c and d). Unlike gtCP, which is expressed in most cells, cjBlue is expressed in a fraction of them, as we see from the corresponding transmission image. Therefore, stimulated emission microscopy opens up the possibility of visualizing chromoproteins that are not accessible by fluorescence microscopy.

Next, we show the stimulated emission imaging in live *E. coli* cells of a non-fluorescent indigo reaction product commonly used to assay gene expression. β -galactosidase catalyses the cleavage of the glycosidic linkage of X-gal, a popular chromogenic substrate, eventually forming an indigo product after oxidation²¹. The gene encoding β -galactosidase, *lacZ*, is a classic reporter gene. Traditionally, the indigo product has to accumulate in sufficient concentration for its blue colour to be visible²¹. With stimulated emission, the accumulation of the indigo product can now be monitored in wild-type cells without induction of *lacZ* (Fig. 3e–g). The inhomogeneous distribution of the indigo product inside individual cells (Fig. 3e and f) is consistent with the fact that the product is insoluble and tends to form localized precipitates. In contrast, the corresponding transmission image (Fig. 3g) shows no signs of any colour. Although an assay

using a fluorogenic substrate has recently been developed²², it requires microfluidics to enclose individual cells because the hydrolysis product there is quickly pumped out by the cell. Hence, stimulated emission microscopy allows us to monitor a product that is assumed to report basal *lacZ* gene activity with superb sensitivity.

Monitoring transdermal delivery of non-fluorescent drug with intrinsic three-dimensional optical sectioning is another application. In particular, we show mapping of a cationic thiazine dye toluidine blue O (TBO) at both the cellular and tissue levels. Having a selective affinity for cancer cells *in vivo*, TBO is a photosensitizer in photodynamic

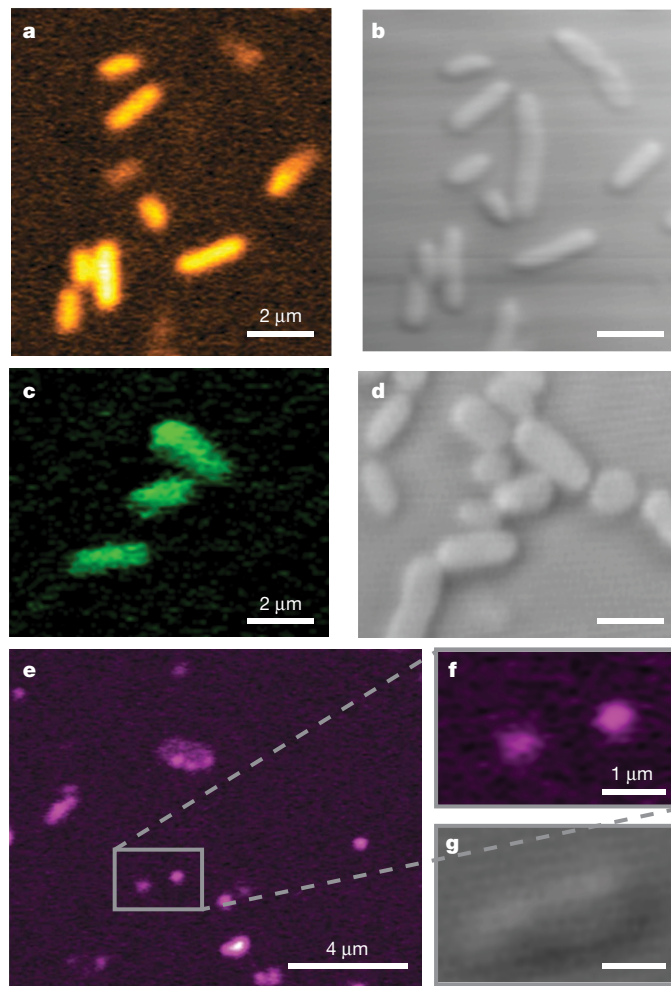


Figure 3 | Imaging non-fluorescent chromoproteins and chromogenic reporter for gene expression. Imaging distributions of cytoplasmic chromoproteins gtCP (**a**) and cjBlue (**c**) in live *E. coli* cells, respectively, by stimulated emission microscopy. **b** and **d** are the corresponding wide-field transmission images. Plasmids containing the genes encoded for gtCP and cjBlue are transformed into *E. coli*, respectively. cjBlue is expressed less abundantly than gtCP inside cells. **e**, Stimulated emission imaging of the oxidation product (4,4'-dichloro-5,5'-dibromoindigo) resulting from X-gal hydrolysis catalysed by β -galactosidase, the protein encoded by the *lacZ* reporter gene. Indigo accumulation in *E. coli* cells is detected without any *lacZ* induction. The signal is detected only when both the cells and X-gal are present. Unlike the homogeneous protein images in **a** and **c**, the indigo product shows inhomogeneous localized distribution inside cells owing to its insolubility, as shown in the magnified image **f**. The direct transmission image **g** shows the morphology of the same region. gtCP and cjBlue absorb around 580 nm (ref. 18) and 610 nm (ref. 19), respectively; indigo product has a broad absorption of 600 to 650 nm. Wavelengths for excitation and stimulation, respectively, are 590 nm and 660 nm in **a** and **c**, and 600 nm and 680 nm in **e** and **f**. The excitation and stimulation beams of 0.1 mW and 0.3 mW, respectively, and time delay of ~ 0.3 ps are used for all the above images. All of the images were taken between 5 s and 45 s, depending on the image size.

therapy^{23,24}. Subcellular localization of TBO is crucial because it influences both the level and the kinetics of apoptosis induction. However, it is difficult to image the true TBO distribution, because its fluorescence is quenched when bound to tissue substrates and only the non-specific stain residue in the tissue retains its native fluorescence²⁴. Independent of the complication from fluorescence contrast, stimulated emission microscopy is suitable for this problem. The stimulated emission image of TBO inside the cancer cells after incubation shows its local accumulation (Supplementary Fig. 1). When topically applied to skin tissue, being hydrophilic and water soluble, TBO is enriched in the centre of the protein phase of the polygonal *stratum corneum* cells rather than in the intercellular space which is in the lipid phase (Fig. 4a). At a level 20 μm deeper, TBO displays a rich subcellular distribution in the cytoplasm of the epidermis, where cellular proliferation actively takes place (Fig. 4b). These imaging results support the hydrophilic delivery pathway as well as the recent hypothesis that TBO binds to cytoplasmic RNA to initiate apoptosis²³.

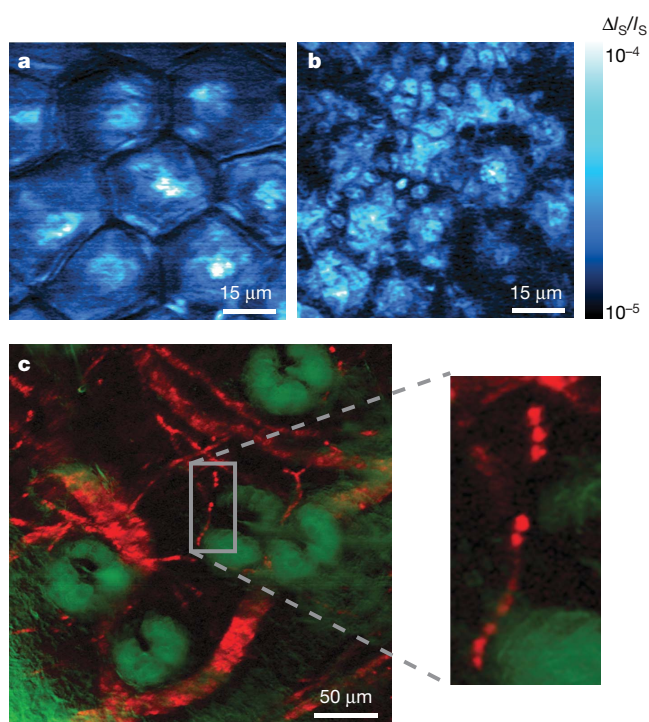


Figure 4 | Transdermal drug distribution in three-dimensional and microvascular imaging. **a, b**, Drug delivery of TBO, a drug used as photosensitizer in photodynamic therapy, to the same area of freshly cut mouse ear skin at two different depths (3 μm and 25 μm), 30 min after topical application of 10 μM TBO/PBS solution. At the surface layer (3 μm) of *stratum corneum*, **a** shows TBO accumulated in the protein phase of the polygonal cells rather than in the lipid-rich intercellular space. At the layer of epidermis (25 μm), **b** shows a rich TBO distribution following the subcellular cytoplasm of nucleated basal keratinocytes. These images support the 'hydrophilic path' as a main pathway for transdermal drug delivery of TBO. Wavelengths are 600 nm and 680 nm for excitation and stimulation, respectively. The excitation and stimulation beams have 0.1 mW and 0.3 mW, respectively, at the focus, and the time delay is ~0.3 ps. The colour table (cyan indicates hot) shows the magnitude of $\Delta I_S/I_S$. **c**, *Ex vivo* imaging of microvasculature network of a mouse ear based on endogenous haemoglobin contrast. The stimulated emission image (red channel, maximum intensity projection) shows the blood vessel network surrounding sebaceous glands (green channel, simultaneously recorded by confocal reflectance). In the zoomed-in image, individual red blood cells are lined up within a single capillary (~5 μm in diameter). In **c**, 830 nm (~20 mW) and 600 nm (~3 mW) are used for two-photon excitation of Soret band and one-photon stimulated emission of Q band of haemoglobin, respectively. Pulse widths of both excitation and stimulation beams are about 0.2 ps with a ~0.2 ps time delay between them.

Finally, we demonstrate label-free imaging of microvascular structure, based on endogenous contrast from non-fluorescent haemoglobin. The structure and haemodynamics of blood vessels play a major role in many biomedical processes, such as angiogenesis in tumours²⁵ and cerebral oxygen delivery in the brain^{26,27}. However, established techniques such as magnetic resonance imaging, computed tomography, positron emission tomography, ultrasound, confocal and two-photon fluorescence microscopy either lack the spatial resolution needed to resolve individual capillaries or require exogenous contrast agents. Here we perform *ex vivo* stimulated emission imaging of the well-developed vascular network from a nude mouse ear, by exciting the Soret band of haemoglobin through efficient two-photon absorption²⁸ and subsequently stimulating the emission from its Q band, which has a longer excited-state lifetime than the Soret band²⁹. As shown in Fig. 4c, blood vessels (in red) branch and loop around sebaceous glands (in green). In addition, single red blood cells can be clearly seen to reside within individual capillaries (~5 μm in diameter). Two-photon absorption has recently been developed to image haemoglobin¹³. We note that stimulated emission is more sensitive because it involves direct one-photon transition. Furthermore, our new technique offers the prospect of three-dimensional mapping of blood oxygenation levels to address a broad range of physiological and pathological problems^{25–27}.

We note that the complexity and cost of the instrument can be much reduced as fibre laser sources are adapted. Stimulated emission microscopy allows imaging of non-fluorescent chromophores with three-dimensional optical sectioning and high sensitivity, and extends the repertoire of reporters for biological imaging beyond fluorophores.

METHODS SUMMARY

In the scheme depicted in Fig. 1, two femtosecond optical parametric oscillators (Coherent/APE) are synchronously pumped by a femtosecond mode-locked 76-MHz Ti:sapphire laser (Coherent). Two independent frequency-doubled outputs from these two optical parametric oscillator signal waves, in the wavelength range of 560 to 700 nm with pulse widths around 200 fs, serve as excitation and stimulation pulse trains. A pulse compressor consisting of a pair of SF11 prisms is built to control the pulse width. Collinear excitation and stimulation beams are combined and focused with a high numerical aperture (NA = 1.2) objective onto a common focal spot. The temporal delay between the synchronized excitation and stimulation inter-pulse is adjusted to between 0.2 and 0.3 ps. The intensity of the excitation beam is modulated by an acousto-optical modulator (Crystal technology) at 5 MHz. A condenser with NA = 0.9 is used to collect the forward propagating stimulation beam, which is spectrally filtered before being detected by a photodiode. To acquire images with a laser scanning microscope (FV300, Olympus), we used a 100- μs time constant for a lock-in amplifier (SR844, Stanford Research) and pixel dwell time of 190 μs .

Received 28 March; accepted 13 August 2009.

- Pawley, J. B. (ed.) *Handbook of Biological Confocal Microscopy* 3rd edn (Springer, 2006).
- Lakowicz, J. R. *Principles of Fluorescence Spectroscopy* (Plenum Press, 1983).
- Turro, N. J. *Modern Molecular Photochemistry* (University Science Books, 1991).
- Einstein, A. On the quantum theory of radiation. *Phys. Z.* **18**, 121 (1917).
- Seigman, A. E. *Laser* 264–307 (University Science Books, 1986).
- Hamilton, C. E., Kinsey, J. L. & Field, R. W. Stimulated emission pumping: new methods in spectroscopy and molecular dynamics. *Annu. Rev. Phys. Chem.* **37**, 493–524 (1986).
- Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
- Dong, C. Y., So, P. T., French, T. & Gratton, E. Fluorescence lifetime imaging by asynchronous pump-probe microscopy. *Biophys. J.* **69**, 2234–2242 (1995).
- Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry* 361–374 (W. H. Freeman, 1980).
- Moerner, W. E. & Kador, L. Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).
- Ye, J., Ma, L. S. & Hall, J. L. Ultrasensitive detections in atomic and molecular physics: demonstration in molecular overtone spectroscopy. *J. Opt. Soc. Am. B* **15**, 6–15 (1998).
- Freudiger, C. W. et al. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* **322**, 1857–1861 (2008).

13. Fu, D. *et al.* High-resolution *in vivo* imaging of blood vessels without labeling. *Opt. Lett.* **32**, 2641–2643 (2007).
14. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
15. Evans, C. L. & Xie, X. S. Coherent anti-Stokes Raman scattering microscopy: chemical imaging for biology and medicine. *Annu. Rev. Anal. Chem.* **1**, 883–909 (2008).
16. Rittweger, E., Rankin, B. R., Westphal, V. & Hell, S. W. Fluorescence depletion mechanisms in super-resolving STED microscopy. *Chem. Phys. Lett.* **442**, 483–487 (2007).
17. Du, H. *et al.* PhotochemCAD: A computer-aided design and research tool in photochemistry. *Photochem. Photobiol.* **68**, 141–142 (1998).
18. Gurskaya, N. G. *et al.* GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett.* **507**, 16–20 (2001).
19. Chan, M. C. Y. *et al.* Structural characterization of a blue chromoprotein and its yellow mutant from the sea anemone *Cnidopus japonicus*. *J. Biol. Chem.* **281**, 37813–37819 (2006).
20. Zhang, J., Campbell, R. E., Ting, A. Y. & Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nature Rev. Mol. Biol.* **3**, 906–918 (2002).
21. Miller, J. H. *Experiments in Molecular Genetics* 171–224 (Cold Spring Harbor Laboratory, 1972).
22. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
23. Tremblay, J. F. *et al.* Photodynamic therapy with toluidine blue in Jurkat cells: cytotoxicity, subcellular localization and apoptosis induction. *Photochem. Photobiol. Sci.* **1**, 852–856 (2002).
24. Chelvanayagam, D. K. & Beazley, L. D. Toluidine blue-O is a Nissl bright-field counterstain for lipophilic fluorescent tracers Di-ASP, Dil and DiO. *J. Neurosci. Methods* **72**, 49–55 (1997).
25. McDonald, D. M. & Choyke, P. L. Imaging of angiogenesis: from microscope to clinic. *Nature Med.* **9**, 713–725 (2003).
26. Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D. & Wiesel, T. N. Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* **324**, 361–364 (1986).
27. Kleinfeld, D., Mitra, P. P., Helmchen, F. & Denk, W. Fluctuations and stimulus-induced changes in blood flow observed in individual capillaries in layers 2 through 4 of rat neocortex. *Proc. Natl Acad. Sci. USA* **95**, 15741–15746 (1998).
28. Clay, G. O., Schaffer, C. B. & Kleinfeld, D. Large two-photon absorptivity of hemoglobin in the infrared range of 780–880 nm. *J. Chem. Phys.* **126**, 025102 (2007).
29. Wang, W. *et al.* Femtosecond multicolor pump-probe spectroscopy of ferrous cytochrome c. *J. Phys. Chem. B* **104**, 10789–10801 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Lukyanov and A. Miyawaki for the gifts of chromoprotein gtCP and cjBlue plasmid DNA, respectively; Coherent Inc. for lending us a femtosecond optical parametric oscillator; and P. Choi for preparing X-gal *E. coli* cells. We also thank B. G. Saar, C. W. Freudiger, S. Basu, J. W. Lichtman and C. B. Schaffer for discussions, and R. Tsien for suggesting the use of chromoproteins. This work was supported by a National Science Foundation (grant CHE-0634788) and the US Department of Energy's Basic Energy Sciences Program (DE-FG02-07ER15875).

Author Contributions W.M., S.L. and S.C. performed experiments and analysed data. R.R. constructed *E. coli* cells expressing chromoproteins. G.R.H. and S.C. helped to construct the laser systems. W.M., S.L. and X.S.X. conceived the concept, designed the experiments and wrote the paper.

Author Information The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to X.S.X. (xie@chemistry.harvard.edu).

LETTERS

Atmospheric carbon dioxide through the Eocene–Oligocene climate transition

Paul N. Pearson¹, Gavin L. Foster² & Bridget S. Wade³

Geological and geochemical evidence^{1–3} indicates that the Antarctic ice sheet formed during the Eocene–Oligocene transition⁴, 33.5–34.0 million years ago. Modelling studies^{5,6} suggest that such ice-sheet formation might have been triggered when atmospheric carbon dioxide levels ($p_{\text{CO}_2}^{\text{atm}}$) fell below a critical threshold of ~ 750 p.p.m.v., but the timing and magnitude of $p_{\text{CO}_2}^{\text{atm}}$ relative to the evolution of the ice sheet has remained unclear. Here we use the boron isotope pH proxy^{7,8} on exceptionally well-preserved carbonate microfossils from a recently discovered geological section in Tanzania^{9,10} to estimate $p_{\text{CO}_2}^{\text{atm}}$ before, during and after the climate transition. Our data suggest that a reduction in $p_{\text{CO}_2}^{\text{atm}}$ occurred before the main phase of ice growth, followed by a sharp recovery to pre-transition values and then a more gradual decline. During maximum ice-sheet growth, $p_{\text{CO}_2}^{\text{atm}}$ was between ~ 450 and $\sim 1,500$ p.p.m.v., with a central estimate of ~ 760 p.p.m.v. The ice cap survived the period of $p_{\text{CO}_2}^{\text{atm}}$ recovery, although possibly with some reduction in its volume, implying (as models predict¹¹) a nonlinear response to climate forcing during melting. Overall, our results confirm the central role of declining $p_{\text{CO}_2}^{\text{atm}}$ in the development of the Antarctic ice sheet (in broad agreement with carbon cycle modelling¹²) and help to constrain mechanisms and feedbacks associated with the Earth's biggest climate switch of the past 65 Myr.

The principal geochemical fingerprint of the Eocene–Oligocene transition (EOT) is an approximately +1.5‰ ‘shift’ towards more positive values of the oxygen isotope ratio of deep-sea carbonates between ~ 34.0 and 33.5 million years (Myr) ago, the last part of which is a prominent ‘step’ of about +0.5‰ at about 33.5 Myr ago^{1,13,14}. The isotope shift corresponds to an extended period of global climatic disruption, widespread extinction and biogeographic reorganization^{4,10}. Recent evidence^{15,16} suggests that the first phase of the isotope shift was caused mainly by global cooling, with a little ice growth, while the end step was a geologically rapid phase of ice growth of less than 40,000 years (< 40 kyr) with a smaller component of deep-sea cooling. The step occurred at a time when Earth's orbital parameters favoured ice-sheet growth^{14,17} and was probably accelerated by positive feedbacks as a massive high-albedo ice cap was assembled^{5,6}. At this point an ‘ice-house’ climate mode characterized by pronounced orbitally paced climate cycles was initiated¹⁷. The early Oligocene world was, nevertheless, warmer than today, with no evidence for sustained continental ice caps in the northern hemisphere, and possibly West Antarctica, until much later^{2,6,18}.

It is crucial for better prediction of future climate change to establish the level of radiative forcing through the EOT^{3,19}. A variety of $p_{\text{CO}_2}^{\text{atm}}$ proxies through the Cenozoic era¹⁹ suggest that values of $> 1,000$ parts per million by volume (p.p.m.v.) were common in the Palaeocene and Eocene epochs, declining to near-modern levels by the early Miocene epoch (~ 23 Myr ago), but there is little agreement in detail between

proxies and few data exist through the critical interval. The most detailed available record is from alkenone-based estimates²⁰, but even in that record there are no data from the EOT.

Here we used boron isotope ($\delta^{11}\text{B}$) analysis of the carbonate shells of upper-ocean planktonic foraminifera to establish palaeo-surface ocean pH (refs 7, 8) from which we infer the dissolved CO_2 concentration, $[\text{CO}_2]_{\text{aq}}$. In oligotrophic locations, $[\text{CO}_2]_{\text{aq}}$ is in approximate equilibrium with $p_{\text{CO}_2}^{\text{atm}}$. The main uncertainties are the value for the boron isotope ratio of seawater ($\delta^{11}\text{B}_{\text{sw}}$), sea surface temperature, and the requirement to estimate one other parameter of the carbonate system (for example, total alkalinity, $[\text{CO}_3^{2-}]$). We used estimates of $\delta^{11}\text{B}_{\text{sw}}$ from boron cycle modelling^{21,22} and a two-box model to estimate $[\text{CO}_3^{2-}]$ changes in both surface and deep waters associated with the substantial stepwise changes in the oceanic calcite compensation depth^{14,23} through the climate transition. Sea surface temperature estimates were made using Mg/Ca measurements on the same shells used for analysing $\delta^{11}\text{B}$, supported by $\delta^{18}\text{O}$ and organic temperature proxy measurements¹⁰.

Geological sections with well-preserved planktonic foraminifera across the climate transition are rare. We recovered a suitable section from Tanzania in boreholes in the Kilwa Group, a bathyal outer shelf/slope deposit with oligotrophic open-ocean conditions^{9,10}. Foraminifer shells are excellently preserved owing to the relatively impermeable clay

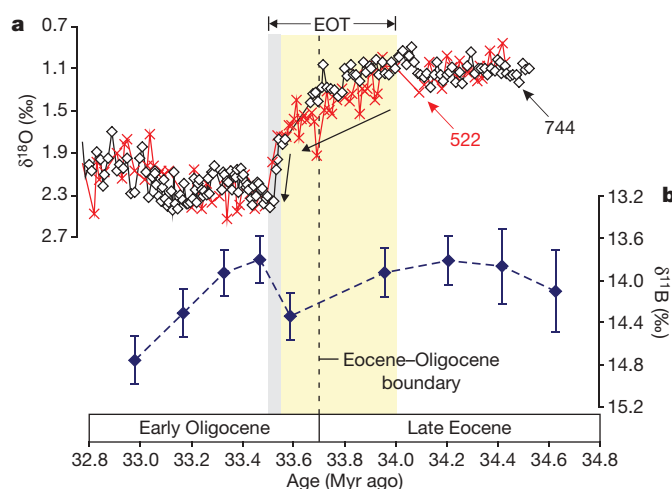


Figure 1 | Deep-sea oxygen isotope records across the EOT compared with boron isotopes from Tanzania. **a**, Deep-sea oxygen isotope records of benthic foraminifera from ODP sites 522 (red crosses) and 744 (black diamonds; ref. 13). Arrows indicate the ‘shift’ and ‘step’ in the isotope records referred to in the text. **b**, Boron isotope ratios of *Turborotalia ampliapertura* (212–250 μm), error bars reflect long-term analytical reproducibility or internal precision, whichever is larger (at 95% confidence).

¹School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3YE, UK. ²Bristol Isotope Group, Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK.

³Department of Geology and Geophysics, Texas A&M University, College Station, Texas 77843-3115, USA.

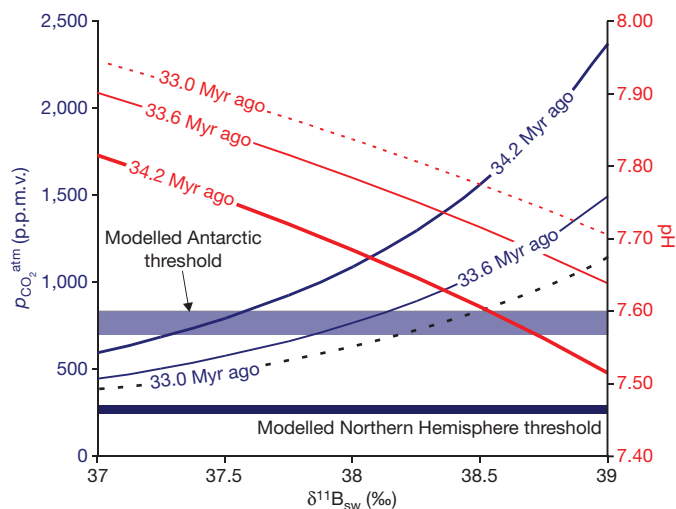


Figure 2 | Sensitivity of pH and atmospheric carbon dioxide estimates to $\delta^{11}\text{B}$ for sea water, with modelled atmospheric p_{CO_2} thresholds for ice growth. Blue lines show reconstructed $p_{\text{CO}_2}^{\text{atm}}$ for three time intervals. Thresholds for glaciation in the northern and southern hemispheres are from refs 5 and 6. Red lines show reconstructed pH values for the three time intervals. pH is reported on the total pH scale.

facies and shallow maximum burial depth of the deposit¹⁰. The section has already produced records of $\delta^{18}\text{O}$ (refs 10, 24) and Mg/Ca (ref. 15) through the EOT. We analysed nine samples spanning a period of ~ 1.6 Myr, with an average sample spacing of ~ 200 kyr. Measurements were made on shells of the non-symbiotic upper-ocean-dwelling foraminifer *Turborotalia ampliapertura*.

All boron isotope results are in the range $\delta^{11}\text{B} = 14.2 \pm 0.5$ (Supplementary Information and Fig. 1). Boron has a residence time in sea water (10–20 Myr)²¹ sufficiently long that $\delta^{11}\text{B}_{\text{sw}}$ is unlikely to have changed significantly within the time window of our study. Because the $\delta^{11}\text{B}$ of foraminifer calcite is positively correlated with pH (refs 7, 8), which is in turn negatively correlated with $[\text{CO}_2]_{\text{aq}}$, the temporal pattern in $\delta^{11}\text{B}$ establishes a qualitative history of $p_{\text{CO}_2}^{\text{atm}}$ evolution regardless of the further assumptions we take in reconstructing absolute values. Our data show a significant increase in $\delta^{11}\text{B}$ in the period of global cooling before the main step in the oxygen isotope record indicating a decline in $p_{\text{CO}_2}^{\text{atm}}$, followed by a recovery to approximately pre-excursion values, and then further subsequent $\delta^{11}\text{B}$ increase (and $p_{\text{CO}_2}^{\text{atm}}$ decline; Fig. 1).

Geochemical models of the boron cycle^{21,22} agree that $\delta^{11}\text{B}_{\text{sw}}$ was probably in the range $+37$ to $+39\text{‰}$ during this time interval. Figure 2 illustrates the sensitivity of our critical pH and $p_{\text{CO}_2}^{\text{atm}}$ estimates before, during and after the EOT to this range in seawater isotopic composition. The central $p_{\text{CO}_2}^{\text{atm}}$ estimate of ~ 760 p.p.m.v. at 33.6 Myr ago (where $\delta^{11}\text{B}_{\text{sw}} = 38.0\text{‰}$) is close to the range identified by ice-sheet models as the ‘threshold’ for Antarctic ice sheet growth (700–840 p.p.m.v.)^{5,6}, but we cannot rule out a value as low as ~ 450 p.p.m.v. or as high as $\sim 1,500$ p.p.m.v.. Even this wide range is useful, because 450 p.p.m.v. is nearly twice the modelled threshold value for Northern Hemisphere continental glaciation⁶, supporting the contention that Oligocene glaciation was most probably unipolar⁶; and values much higher than 1,500 p.p.m.v. have been used in models of late Eocene climate²⁵. Improvements in reconstructing $\delta^{11}\text{B}_{\text{sw}}$ are clearly required to refine our absolute estimates of $p_{\text{CO}_2}^{\text{atm}}$.

Using the central estimate in which $\delta^{11}\text{B}_{\text{sw}} = 38.0\text{‰}$ (ref. 21), the data indicate a significant increase in pH from ~ 7.7 to ~ 7.8 across

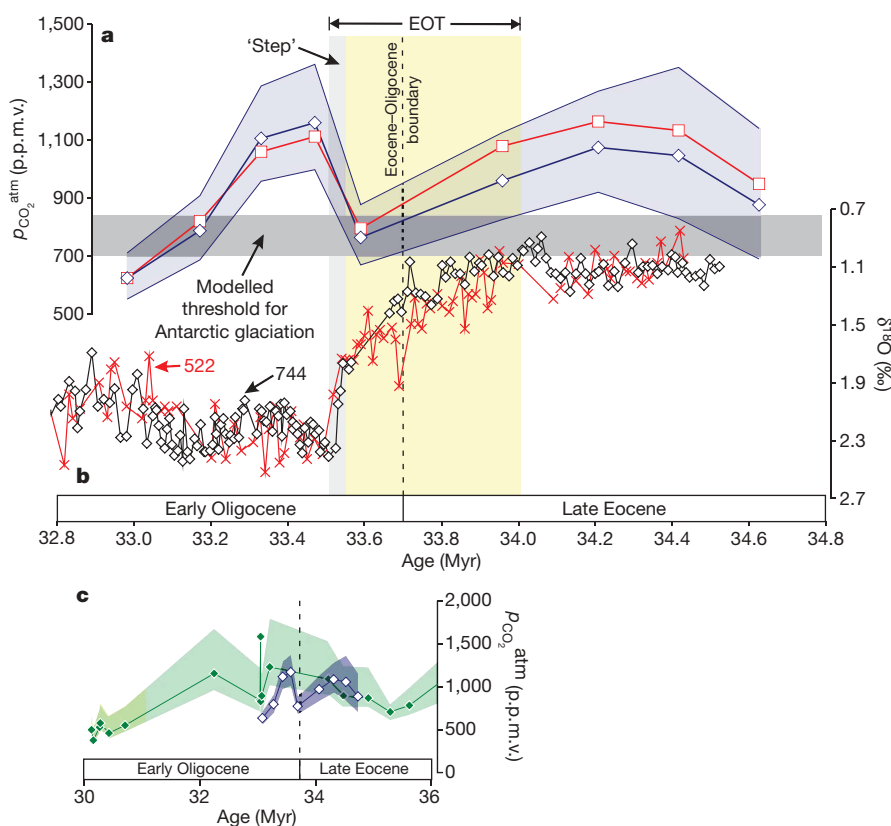


Figure 3 | Reconstructed $p_{\text{CO}_2}^{\text{atm}}$ compared with the deep-sea benthic foraminiferal stable isotope record. **a**, $p_{\text{CO}_2}^{\text{atm}}$ (\pm uncertainty from $\delta^{11}\text{B}$ measurements at 95% confidence). Blue symbols are calculated using modelled $[\text{CO}_3^{2-}]$ changes based on varying calcite compensation depth; red symbols are for constant $[\text{CO}_3^{2-}]$. The grey band is the threshold for

Antarctic glaciation^{5,6}. **b**, Deep-sea oxygen isotopes from DSDP site 522 (red crosses) and ODP site 744 (black diamonds; ref. 13). **c**, Alkenone proxy estimates (green symbols with green shaded band for maximum and minimum; ref. 20) and $\delta^{11}\text{B}$ proxy measurements (this study; colouring as in **a**).

the EOT that equates to a reduction in $p_{\text{CO}_2}^{\text{atm}}$ from $\sim 1,100$ p.p.m.v. to ~ 760 p.p.m.v. over the extended shift in the $\delta^{18}\text{O}$ curve, albeit with a large uncertainty (Fig. 3). A significant rise in $p_{\text{CO}_2}^{\text{atm}}$ to $\sim 1,150$ p.p.m.v. occurred after the main step in the $\delta^{18}\text{O}$ record, which was followed by further reductions to ~ 625 p.p.m.v. over the next 500 kyr. These $p_{\text{CO}_2}^{\text{atm}}$ estimates are in broad agreement with the alkenone proxy-derived estimates²⁰ (Fig. 3c) but add more detail, including the inferred $p_{\text{CO}_2}^{\text{atm}}$ decline before the $\delta^{18}\text{O}$ step and the sharp increase afterwards.

Cooling in the early phases of the EOT was associated with some ice growth^{15,16}, which models^{5,6} and geomorphological observations²⁶ suggest was probably in isolated regions of Antarctica. The decline in $p_{\text{CO}_2}^{\text{atm}}$ detected in our study may have contributed to global cooling and preconditioned the system for explosive ice sheet growth at the time of the $\delta^{18}\text{O}$ step (~ 33.5 Myr ago) when orbital parameters were favourable^{5,14}. The data indicate that once established, the Antarctic ice sheet withstood an increase in $p_{\text{CO}_2}^{\text{atm}}$ to levels that were similar to or slightly higher than before the climate transition. This hysteresis effect¹¹ occurs because the bright surface of the ice cap reflects more sunlight, and once an ice cap has formed, melting at its margin is compensated by flow from the cold, high altitude interior. Details in the $\delta^{18}\text{O}$ record from ODP site 744 (Fig. 3) suggest a slight rebound towards more negative values between about 33.5 and 33.3 Myr ago that could indicate partial melt-back of the ice cap and/or warming during this time of relatively high $p_{\text{CO}_2}^{\text{atm}}$. However, we note that this feature is not obvious in all $\delta^{18}\text{O}$ records through this interval^{13,14}.

Recent carbon cycle modelling¹² has used constraints provided by the stepwise fall in the calcite compensation depth^{14,23} and observed deep-sea carbon isotope excursions^{13,14} to explore a variety of possible feedbacks in the global carbon cycle during the EOT. The model preferred in ref. 12 involves a permanent increase in the ratio of calcium carbonate buried in deep-water environments compared with shallow seas as a consequence of the glacioeustatic sea-level fall. This scenario predicts a short-lived $p_{\text{CO}_2}^{\text{atm}}$ drop of ~ 300 p.p.m.v. over the transition, followed by a rebound to levels slightly higher than before the event, which is similar to the $p_{\text{CO}_2}^{\text{atm}}$ evolution suggested by our data. However, our study indicates that $p_{\text{CO}_2}^{\text{atm}}$ recovered more rapidly (within ~ 50 kyr) after the step in the $\delta^{18}\text{O}$ record than in the most favoured model of ref. 12 (~ 500 kyr). Moreover, the initial $p_{\text{CO}_2}^{\text{atm}}$ decline that we observe is associated with the long shift in the $\delta^{18}\text{O}$ record, which appears to have had only a relatively small component of ice growth^{15,16}, and hence would not have been associated with the maximum sea-level fall. A combination of several mechanisms, some of which are explored in ref. 12, is therefore needed to fit the additional constraint provided by our $p_{\text{CO}_2}^{\text{atm}}$ record. This may include release of organic carbon from geological reservoirs, changes in ocean productivity and circulation patterns, and variations in rock weathering rates^{12,27,28}.

The possible roles of declining greenhouse-gas forcing and tectonic gateway effects in promoting Antarctic glaciation across the Eocene–Oligocene boundary have been debated for decades^{1,2,4}. Recent studies point to a consensus that ice growth followed an extended period of global cooling across continents^{18,26} and in the oceans in shallow^{15,16} and deep^{13,14} water and in high^{18,25,26} and low^{15,16,25} latitudes. This strongly suggests that the primary cause was a diminishing greenhouse effect²⁵. Although greenhouse gases other than CO_2 (for which there are no proxies) may have contributed, changing $p_{\text{CO}_2}^{\text{atm}}$ is likely to have had the greatest forcing¹⁹. Ours is the first proxy-based study to confirm a substantial $p_{\text{CO}_2}^{\text{atm}}$ decline during the climate transition. We also find a sharp $p_{\text{CO}_2}^{\text{atm}}$ increase after maximum ice growth as the global carbon cycle adjusted to the presence of a large ice cap and there was a nonlinear hysteresis effect as the ice cap withstood this transient $p_{\text{CO}_2}^{\text{atm}}$ rise. This study reaffirms the links between cryosphere development and atmospheric carbon dioxide levels at the largest and most important climatic tipping point of the last 65 million years.

METHODS SUMMARY

Foraminiferal samples were separated from 50-cm core segments from Tanzania Drilling Project sites 12 and 17 (ref. 9). The composite age model for these sites¹⁰ is on the timescale of ref. 29. Foraminiferal calcite is exceptionally well-preserved and between 300 and 1,000 tests of *Turborotalia ampliapertura* (212–250 μm) were separated for each boron isotope analysis. Trace element and boron isotope analyses were carried out following ref. 8. Trace element and isotopic data are listed in the Supplementary Information. The pH is calculated from $\delta^{11}\text{B}$ using tandem Mg/Ca-derived temperatures and following ref. 8.

The second parameter of the carbonate system required to estimate $p_{\text{CO}_2}^{\text{atm}}$ is calculated for each sample from the evolution of the calcite compensation depth over the EOT, using a simple two-box model and CaCO_3 mass accumulation rate in the equatorial Pacific^{14,23}. In the deep box (deep ocean) at the saturation depth (assumed to be approximated by the calcite compensation depth), $[\text{CO}_3^{2-}]$ equals $[\text{CO}_3^{2-}]$ at saturation. We assume a constant $[\text{CO}_3^{2-}]$ gradient between the deep box and upper box (surface ocean) of $\sim 72 \mu\text{mol mol}^{-1}$ such that at 32.98 Myr ago (post-EOT) the surface water is 6.5 times oversaturated, the same as the pre-industrial western Indian Ocean³⁰. By combining our pH estimate from $\delta^{11}\text{B}$ with this estimate of $[\text{CO}_3^{2-}]_{\text{surface}}$ we can calculate $[\text{CO}_2]_{\text{aq}}$ and thus $p_{\text{CO}_2}^{\text{atm}}$ using Henry's Law. Changing this gradient by $\pm 25\%$ or changing the depth of the calcite compensation depth by ± 500 m does not alter our $p_{\text{CO}_2}^{\text{atm}}$ record beyond the uncertainty imposed by the measurement uncertainty in $\delta^{11}\text{B}$. Keeping $[\text{CO}_3^{2-}]_{\text{surface}}$ constant at post-EOT values also does not greatly influence our $p_{\text{CO}_2}^{\text{atm}}$ record (Fig. 3). This implies that our assumptions regarding this second parameter of the carbonate system are of secondary importance and the $p_{\text{CO}_2}^{\text{atm}}$ evolution we reconstruct is driven predominantly by the measured $\delta^{11}\text{B}$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 May; accepted 21 August 2009.

Published online 13 September 2009.

- Kennett, J. P. & Shackleton, N. J. Oxygen isotopic evidence for the initiation of the psychrosphere 38 Myr ago. *Nature* **260**, 513–515 (1976).
- Barrett, P. in *Developments in Earth and Environmental Sciences* (eds Florindo, F. & Siebert, M.) Vol. 8, 34–83 (Elsevier, 2009).
- Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
- Coxall, H. K. & Pearson, P. N. in *Deep-time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies* (eds Williams, M. et al.) 351–387 (The Micropalaeontological Society, Special Publications, The Geological Society, London, 2007).
- Deconto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica triggered by declining atmospheric CO_2 . *Nature* **421**, 245–249 (2003).
- Deconto, R. M. et al. Thresholds for Cenozoic bipolar glaciation. *Nature* **455**, 652–656 (2008).
- Hemming, N. G. & Hanson, G. N. Boron isotope composition and concentration in modern marine carbonates. *Geochim. Cosmochim. Acta* **59**, 371–379 (1992).
- Foster, G. L. Seawater pH, $p\text{CO}_2$ and $[\text{CO}_3^{2-}]$ variations in the Caribbean Sea over the last 130 kyr: a boron isotope and B/Ca study of planktic foraminifera. *Earth Planet. Sci. Lett.* **271**, 254–266 (2008).
- Nicholas, C. J. et al. Stratigraphy and sedimentology of the Upper Cretaceous to Paleogene Kilwa Group, southern coastal Tanzania. *J. Afr. Earth Sci.* **45**, 431–466 (2006).
- Pearson, P. N. et al. Extinction and environmental change across the Eocene–Oligocene boundary in Tanzania. *Geology* **36**, 179–182 (2008).
- Pollard, D. & Deconto, R. M. Hysteresis in Cenozoic Antarctic ice-sheet variations. *Global Planet. Change* **45**, 9–21 (2005).
- Merico, A., Tyrell, T. & Wilson, P. A. Eocene/Oligocene ocean de-acidification linked to Antarctic sea-level fall. *Nature* **452**, 979–983 (2008).
- Zachos, J. C., Quinn, T. M. & Salamy, K. A. High resolution (10^4 years) deep-sea foraminiferal stable isotope records of the Eocene–Oligocene climate transition. *Paleoceanography* **11**, 251–256 (1996).
- Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2003).
- Lear, C. H., Bailey, T. R., Pearson, P. N., Coxall, H. K. & Rosenthal, Y. Cooling and ice growth across the Eocene–Oligocene transition. *Geology* **36**, 251–254 (2008).
- Katz, M. E. et al. Stepwise transition from the Eocene greenhouse to the Oligocene icehouse. *Nature Geosci.* **1**, 329–334 (2008).
- Pälike, H. et al. The heartbeat of the Oligocene climate system. *Science* **314**, 1894–1898 (2006).
- Eldrett, J. S., Greenwood, D. R., Harding, I. C. & Huber, M. Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. *Nature* **459**, 969–974 (2009).
- Hansen, J. et al. Target atmospheric CO_2 : Where should humanity aim? *Open Atmos. Sci. J.* **2**, 217–231 (2008).

20. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. M. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).
21. Lemarchand, D., Gaillardet, J., Lewin, É. & Allègre, C. J. Boron isotope systematics in large rivers: implications for the marine boron budget and paleo-pH reconstruction over the Cenozoic. *Chem. Geol.* **190**, 123–140 (2002).
22. Simon, L., Lecuyer, C., Marechal, C. & Coltice, N. Modelling the geochemical cycle of boron: implications for the long-term $\delta^{11}\text{B}$ evolution of seawater and oceanic crust. *Chem. Geol.* **225**, 61–76 (2006).
23. Rea, D. K. & Lyle, M. W. Paleogene calcite compensation depth in the eastern subtropical Pacific: answers and questions. *Paleoceanography* **20**, PA1012, doi:10.1029/2004PA001064 (2005).
24. Wade, B. S. & Pearson, P. N. Planktonic foraminiferal turnover, diversity fluctuations and geochemical signals across the Eocene/Oligocene boundary in Tanzania. *Mar. Micropal.* **68**, 244–255 (2008).
25. Liu, Z. *et al.* Global cooling during the Eocene–Oligocene climate transition. *Science* **323**, 1187–1190 (2009).
26. Bo, S. *et al.* The Gamburtsev mountains and the origin and early evolution of the Antarctic ice sheet. *Nature* **459**, 690–693 (2009).
27. Salamy, K. A. & Zachos, J. C. Latest Eocene–early Oligocene climate change and Southern Ocean fertility: inferences from sediment accumulation and stable isotope data. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **145**, 61–77 (1999).
28. Zachos, J. C. & Kump, L. R. Carbon cycle feedbacks and the initiation of Antarctic glaciation in the earliest Oligocene. *Glob. Planet. Change* **47**, 51–66 (2005).
29. Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity timescale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).
30. Key, R. M. *et al.* A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* **18**, GB4031, doi:10.1029/2004GB002247 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NERC grants to P.N.P., B.S.W. and G.L.F. We thank the Tanzania Petroleum Development Corporation, the Tanzania Commission for Science and Technology and the Tanzania Drilling Project field team for support. We are grateful to T. Elliott for discussions.

Author Contributions P.N.P. led the study and fieldwork, prepared foraminifer samples for isotope analysis and wrote the initial draft. G.L.F. conducted all isotope and trace element analyses and calculations and drafted the figures. B.S.W. contributed to fieldwork and prepared foraminifer samples for trace element analyses. All authors contributed to the final text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.N.P. (pearsonp@cardiff.ac.uk).

METHODS

Sample preparation. Foraminiferal tests with extensive visible pyrite contamination were separated before crushing. All samples were crushed and cleaned following the method in ref. 31 except that we did not carry out a reductive step and samples were dissolved in 0.1 M HNO₃ with centrifuging shortly after dissolution to remove any insoluble residue. Analysis of the residue after dissolution and repeat dissolutions of the separated pyrite-rich subsamples indicate that pyrite does not contain significant Mg or B and the samples can be considered free of pyrite contamination, provided the measured Fe/Ca was <3,000 $\mu\text{mol mol}^{-1}$. The majority of samples here have Fe/Ca <1,000 $\mu\text{mol mol}^{-1}$.

Trace element analysis and temperature estimates. The samples of *Dentoglobigerina* cf. *D. tapuriensis*, *Hantkenina* spp. and a small aliquot (10%) of the *T. ampliapertura* selected for isotope analysis were analysed for trace element composition using a ThermoFisher Scientific Element 2 at the University of Bristol following the methodology outlined in ref. 8. Trace element data are tabulated in Supplementary Table 2. This tandem trace element analysis allowed an assessment of cleaning efficiency and Mg/Ca ratios required to estimate temperature (Supplementary Table 2). The long-term reproducibility of our Mg/Ca ratios is around 2% at 95% confidence, based on repeat measurements of three consistency standards over the duration of this study. Sea water temperatures were calculated using a generic Mg/Ca temperature equation³² and assuming a seawater Mg/Ca ratio of 4.3 mol mol⁻¹ (ref. 33). Sea surface temperatures are ~2 °C warmer than the temperature we calculate here for *T. ampliapertura*^{15,24}. Our temperature evolution across the EOT is therefore in good agreement with a number of recent studies^{15,16,25}. The similarity between *Hantkenina* spp. and *T. ampliapertura* support our assignment of the latter as an upper-ocean dweller (originally based on $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ measurements²⁴). The Mg/Ca ratios of *Dentoglobigerina* cf. *D. tapuriensis* show little change across the EOT, and in all cases this species records lower temperatures owing to its deeper habitat²⁴.

Boron isotope analysis and palaeo-pH. Boron was separated from the dissolved samples of *T. ampliapertura* using boron-specific anion exchange resin⁸. The boron isotopic composition was determined using a sample-standard bracketing routine on a ThermoFisher Scientific Neptune multicollector inductively coupled plasma mass spectrometer (MC-ICPMS) at the University of Bristol⁸. The full boron isotope data set is shown in Supplementary Table 1 and our long-term reproducibility is $\pm 0.22\text{‰}$ at 95% confidence, based on repeat total procedural replicates of similar-sized samples of our in-house carbonate standards ($\delta^{11}\text{B} = 24.67 \pm 0.22\text{‰}$; 2 s.d., $n = 11$ and $\delta^{11}\text{B} = 23.93 \pm 0.22\text{‰}$; 2 s.d., $n = 18$).

When measured by MC-ICPMS, and using the isotopic fractionation factor of ref. 34, the $\delta^{11}\text{B}$ of planktonic foraminiferal carbonate is very close (<0.8‰) to the $\delta^{11}\text{B}$ of $\text{B}(\text{OH})_4^-$ ion in sea water⁸, which is itself pH-dependent⁷. The following equation, modified from ref. 7, can thus be used to estimate pH on the total scale:

$$\text{pH} = \text{pK}_{\text{boric}}^* - \log \left[-\frac{\delta^{11}\text{B}_{\text{sw}} - \delta^{11}\text{B}_{\text{foram}}}{\delta^{11}\text{B}_{\text{sw}} - (1.0272\delta^{11}\text{B}_{\text{foram}}) - 27.2} \right]$$

where $\text{pK}_{\text{boric}}^*$ is $-\log_{10}$ of the stoichiometric equilibrium constant for boric acid³⁵ at the *in situ* temperature and salinity (from Mg/Ca temperature and assumed to be 35 practical salinity units, p.s.u., respectively), $\delta^{11}\text{B}_{\text{sw}}$ is the isotopic composition of sea water ($\delta^{11}\text{B} = 37\text{--}39\text{‰}$), and $\delta^{11}\text{B}_{\text{foram}}$ is the measured isotopic composition of the foraminifer *T. ampliapertura*. In studies of modern planktonic foraminifera intra-species $\delta^{11}\text{B}$ differences are largely due to different depth habitats, and small offsets from the $\delta^{11}\text{B}$ of $\text{B}(\text{OH})_4^-$ relate to symbiont activity and foraminiferal respiration (ref. 8 and references therein). *T. ampliapertura* is an extinct surface-dwelling non-spinose foraminifer so here we assume that there is no offset between the measured $\delta^{11}\text{B}$ of *T. ampliapertura* and $\text{B}(\text{OH})_4^-$ in sea water and that the calculated pH is that of the mixed layer. This assumption does not affect our reconstructions unduly because it is of secondary magnitude given the range of $\delta^{11}\text{B}$ of sea water we use and is likely to be constant between samples. Uncertainties in pH for a given $\delta^{11}\text{B}_{\text{sw}}$ are

dominated by measurement uncertainties in the $\delta^{11}\text{B}$. We note that our $\delta^{11}\text{B}$ data are not directly comparable with earlier Cenozoic boron isotope measurements³⁶ because significant inter-laboratory bias exists⁸.

The second carbonate system parameter. To estimate $p_{\text{CO}_2}^{\text{atm}}$ from $\delta^{11}\text{B}$ -derived pH, a second parameter of the carbonate system is needed. We use a simple two-box model to accomplish this. The $[\text{Ca}]_{\text{sw}}$ concentration of the Eocene ocean has been estimated from fluid inclusion studies to be in the range 12 to 20 mmol kg⁻¹ (ref. 37) and following previous studies we use the central value of 17 mmol kg⁻¹ (refs 37, 38). At saturation $[\text{CO}_3^{2-}] = K_{\text{sp}}/[\text{Ca}]$, where K_{sp} is the solubility product of CaCO_3 adjusted for the influence of changing $[\text{Mg}]$ and $[\text{Ca}]$ of sea water (following ref. 38). In the lower box of our model (which represents the deep ocean) we alter $[\text{CO}_3^{2-}]$ until the saturation horizon (the depth separating waters oversaturated with respect to CaCO_3 from waters undersaturated below) is at 4 km—a depth that is approximately that of the calcium carbonate compensation depth in the Pacific Ocean after the EOT²³. The $[\text{CO}_3^{2-}]$ is then altered through the EOT, guided by the CaCO_3 mass accumulation rate at ODP site 1218 (ref. 14). Working back through time we initially shoal the calcite compensation depth to 3.5 km, then deepen it to 4.5 km coincident with the maximum in $\delta^{18}\text{O}$ and CaCO_3 mass accumulation rate at ODP site 1218 (refs 14, 23). Then, in two steps we decrease the depth to 2.5 km, reflecting the minimum in CaCO_3 mass accumulation rate at ODP site 1218 at ~34.0 Myr ago. We then deepen and hold the calcite compensation depth at 3 km until ~34.6 Myr ago. To determine the $[\text{CO}_3^{2-}]$ in the upper box (the surface ocean) we assume a constant gradient in $[\text{CO}_3^{2-}]$ of ~72 $\mu\text{mol mol}^{-1}$ between the two boxes. The magnitude of this gradient is fixed so that at 32.98 Myr ago the degree of CaCO_3 oversaturation in the surface box is similar to the pre-industrial value in the western Indian Ocean (6.5 times oversaturated; calculated from ref. 30).

By combining our pH estimate from $\delta^{11}\text{B}$ with this estimate of $[\text{CO}_3^{2-}]$ we can calculate $[\text{CO}_2]_{\text{aq}}$ and thus $p_{\text{CO}_2}^{\text{atm}}$ using Henry's Law. For the detail of these calculations see ref. 8 and references therein. Following ref. 38 we also account in these calculations for the effect of changing $[\text{Mg}]$ and $[\text{Ca}]$ of sea water on the carbonic acid dissociation constants K_1 and K_2 . Although large changes in our assumed $[\text{CO}_3^{2-}]$ gradient are unlikely³⁸, we note that changing its magnitude by $\pm 25\%$ or changing the depth of the calcite compensation depth by ± 500 m does not alter our $p_{\text{CO}_2}^{\text{atm}}$ record beyond the uncertainty imposed by the measurement uncertainty in $\delta^{11}\text{B}$ for a given $\delta^{11}\text{B}_{\text{sw}}$. Also, although a temperature and salinity reconstruction is required to reconstruct pH, the resultant record is relatively insensitive to inaccuracies in these parameters. For example, a 2 °C temperature change alters $p_{\text{CO}_2}^{\text{atm}}$ by <50 p.p.m.v. and a 2 p.s.u. salinity change alters $p_{\text{CO}_2}^{\text{atm}}$ by <10 p.p.m.v.

- Barker, S., Greaves, M. & Elderfield, H. A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry. *Geochim. Geophys. Geosyst.* **4**, 8407, doi:10.1029/2003GC000559 (2003).
- Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18**, 1050, doi:10.1029/2002PA000846 (2003).
- Wilkinson, B. H. & Algeo, T. J. Sedimentary carbonate record of calcium-magnesium cycling at the Earth's surface. *Am. J. Sci.* **289**, 1158–1194 (1989).
- Klochko, K., Kaufman, A. J., Yao, W., Berne, R. H. & Tossell, J. A. Experimental measurement of boron isotope fractionation in seawater. *Earth Planet. Sci. Lett.* **248**, 276–285 (2006).
- Dickson, A. G. Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K. *Deep-Sea Res.* **37**, 755–766 (1990).
- Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).
- Horita, J., Zimmermann, H. & Holland, H. D. Chemical evolution of seawater during the Phanerozoic: implications from the record of marine evaporates. *Geochim. Cosmochim. Acta* **66**, 3733–3756 (2002).
- Tyrrell, T. & Zeebe, R. E. History of carbonate ion concentration over the last 100 million years. *Geochim. Cosmochim. Acta* **68**, 3521–3534 (2004).

LETTERS

Trench-parallel anisotropy produced by serpentine deformation in the hydrated mantle wedge

Ikuo Katayama¹, Ken-ichi Hirauchi¹, Katsuyoshi Michibayashi² & Jun-ichi Ando¹

Seismic anisotropy is a powerful tool for detecting the geometry and style of deformation in the Earth's interior, as it primarily reflects the deformation-induced preferred orientation of anisotropic crystals^{1,2}. Although seismic anisotropy in the upper mantle is generally attributed to the crystal-preferred orientation of olivine³, the strong trench-parallel anisotropy (delay time of one to two seconds) observed in several subduction systems^{4,5} is difficult to explain in terms of olivine anisotropy, even if the entire mantle wedge were to act as an anisotropic source. Here we show that the crystal-preferred orientation of serpentine, the main hydrous mineral in the upper mantle, can produce the strong trench-parallel seismic anisotropy observed in subduction systems. High-pressure deformation experiments reveal that the serpentine *c*-axis tends to rotate to an orientation normal to the shear plane during deformation; consequently, seismic velocity propagating normal to the shear plane (plate interface) is much slower than that in other directions. The seismic anisotropy estimated for deformed serpentine aggregates is an order of magnitude greater than that for olivine⁶, and therefore the alignment of serpentine in the hydrated mantle wedge results in a strong trench-parallel seismic anisotropy in the case of a steeply subducting slab. This hypothesis is also consistent with the presence of a hydrous phase in the mantle wedge, as inferred from anomalously low seismic-wave velocities⁷.

The geometry and strength of seismic anisotropy are generally estimated by measuring shear-wave splitting, which constrains the polarization direction of the fast shear-wave propagation and the delay time between the arrivals of the fast and slow shear-waves. Measurements of shear-wave splitting above the mantle wedge in most subduction zones, including the Japan subduction system, reveal trench-parallel anisotropy⁸, in which the propagation direction of the fast shear-wave is oriented subparallel to the trench. However, the delay time shows marked spatial variations (Fig. 1a). For example, delay time is ~0.1–0.2 s in northeast Japan⁹ and ~1–2 s beneath the Ryukyu arc⁴. We focus here on splitting data collected from local slab earthquakes, because these provide direct constraints on the seismic anisotropy of the mantle wedge.

The length of the delay time depends on the strength of the anisotropy and thickness of the anisotropic layer. If we consider that seismic anisotropy is caused by the crystal-preferred orientation of olivine, the short delay time recorded in northeast Japan could be explained in terms of a relatively thin anisotropic layer in the mantle wedge¹⁰, whereas the longer delay time observed in the Ryukyu arc cannot be explained in terms of olivine anisotropy, because the inferred anisotropic layer (~100–200 km) would be thicker than the entire mantle wedge sampled by the local shear wave (Fig. 1b).

Serpentine has a strong crystallographic anisotropy of shear-wave velocity (A_V of ~82% for lizardite single-crystal¹¹), approximately five times stronger than that for olivine (A_V of ~18%)¹², and the occurrence

of serpentine is expected where water infiltrates into the mantle¹³. Accordingly, one possible explanation of the strong anisotropy observed beneath the Ryukyu arc is the crystal-preferred orientation

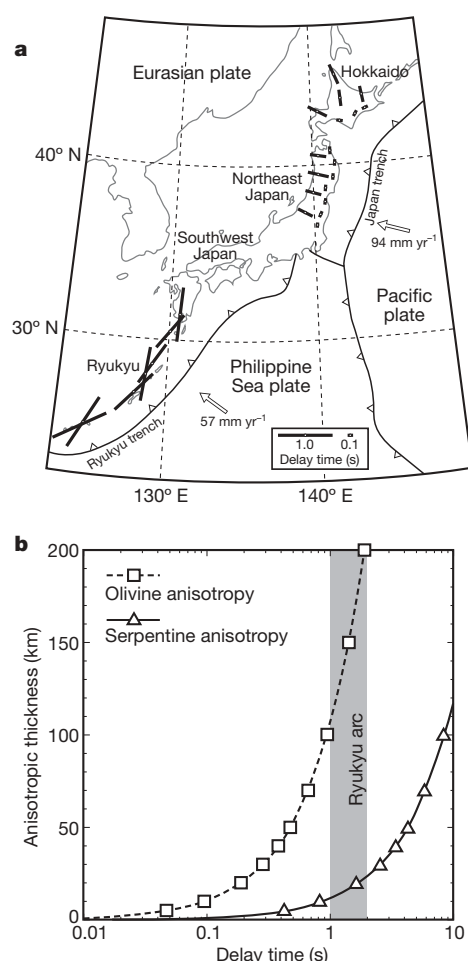


Figure 1 | Observations of shear-wave splitting in the Japan subduction system. **a**, Shear-wave splitting from local S-phases along the Ryukyu arc and northeast Japan^{9,26}. The orientations of bars correspond to the fast polarization direction, and the bar lengths are normalized to the average delay time. **b**, Relation between shear-wave delay time and required thickness of anisotropic layer, assuming olivine and serpentine as a source. If we use the strength of anisotropy based on experimentally deformed B-type olivine aggregates (A_V of 4.5%)⁶ or highly deformed serpentine (this study; A_V of 32%), the long delay time (1–2 s) observed in the Ryukyu arc requires an anisotropic layer of thickness 100–200 km in the case of olivine, but only 10–20 km thickness in the case of serpentine.

¹Department of Earth and Planetary Systems Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan. ²Institute of Geosciences, Shizuoka University, Shizuoka 422-8529, Japan.

of serpentine¹⁴. To test this hypothesis, we performed deformation experiments on serpentine under conditions corresponding to the mantle wedge, and analysed the resulting deformation-induced crystallographic preferred orientation of serpentine in terms of its ability to explain the strong trench-parallel seismic anisotropy.

Deformation experiments were carried out at a pressure of 1 GPa and temperature of 300–400 °C at a constant strain rate using a solid-medium apparatus. The starting material of natural antigorite (a high-temperature form of serpentine), which exhibited near-isotropic texture, was sandwiched between alumina pistons cut at 45° from the maximum compression direction, oriented to represent the deformation geometry of simple shear. We made tiny grooves at the interface between the sample and piston to prevent slip during deformation. These materials were surrounded by a nickel jacket, and the oxygen fugacity was buffered by the Ni/NiO reaction. When the pressure and temperature had reached the desired values, a piston was advanced at a constant rate. Shear strain was measured from the rotation of a nickel strain-marker that was initially orientated perpendicular to the shear direction.

The serpentine samples were deformed up to strains of $\gamma \approx 2.0$ in a dominantly simple-shear geometry (compressional strain was generally less than 10% of shear strain). The serpentine grains and trails of magnetite became significantly elongated, with stretching axes oriented subparallel to the strain-marker (Fig. 2). The deformed serpentine showed undulatory extinction, suggesting intracrystalline deformation via dislocation creep, consistent with the results of mechanical creep tests of serpentine¹⁵.

Figure 3 shows the crystallographic orientations of the deformed serpentine grains, as measured using electron backscatter diffraction (EBSD) at the Centre for Instrumental Analysis, Shizuoka University,

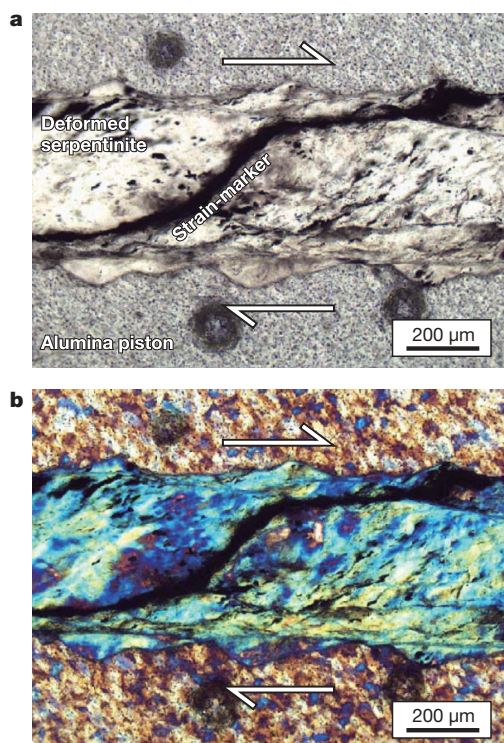


Figure 2 | Optical microphotographs of experimentally deformed serpentine. **a**, Viewed in plane-polarized light. **b**, Viewed with the quartz λ plate inserted. The sample is sandwiched between alumina pistons, and the thin dark layer is a nickel strain-marker, which was originally oriented normal to the shear plane. Arrows indicate the sense of shear. The sample consists mainly of antigorite, with minor magnetite (opaque grains). The serpentine grains and trails of magnetite are strongly elongated subparallel to the strain-marker, and the highly strained grains of serpentine show undulatory extinction.

Japan. Although the starting materials showed at most only a weak crystal-preferred orientation (Fig. 3a), the deformed serpentines in all experiments are characterized by a strong concentration of [001] axes normal to the shear plane and a relatively weak concentration of [100] axes subparallel to the shear direction (Fig. 3b, c). At higher strains, serpentine [100] axes are oriented close to the shear direction, whereas at lower strains the [100] maxima are tilted towards the finite strain ellipse (Fig. 3b). The strong alignment of serpentine (001) planes parallel to the shear plane is consistent with the pattern found in naturally deformed serpentinites¹⁶.

The serpentine fabric produced by the deformation experiments corresponds to seismic anisotropy characterized by a slow propagation direction oriented normal to the shear plane, because the layered crystallographic structure within serpentine (antigorite) results in a much slower seismic velocity propagating perpendicular to the basal (001) plane. Although there is a lack of elastic constants for antigorite because of the complexity of its crystal structure, as a first approximation we calculated the seismic anisotropy corresponding to the deformed serpentine aggregates using the elastic constants for lizardite serpentine¹¹, which has a similar crystal structure and elastic properties to those of antigorite. The results reveal a strong seismic anisotropy for the deformed serpentine aggregates, up to 46% of compressional (P)-wave azimuthal anisotropy and up to 32% of shear (S)-wave anisotropy for the highly sheared sample ($\gamma \approx 2.0$; Fig. 3e). These estimates are similar to the results of elastic wave measurements of naturally deformed serpentinites, which reveal a strong anisotropy in the direction normal to foliation (up to A_V of $\sim 40\%$)^{16,17}. Kern *et al.*¹⁶ demonstrated that the magnitude of anisotropy in serpentinite is relatively insensitive to pressure and temperature in the range of <600 MPa and <600 °C, although the anisotropy may decrease at higher pressures owing to efficient compressibility along the serpentine [001] axis.

We now apply the above results in explaining seismic anisotropy within a subduction zone setting. In the Ryukyu arc, where the relatively young Philippine Sea plate is subducting beneath the Eurasia plate at a steep angle ($\sim 45^\circ$)¹⁸, water is released by dehydration reactions in the subducting crustal material, leading to the formation of serpentine (the main hydrous mineral in the upper mantle) in the overlying mantle wedge. Deformation is primarily induced by the movement of the subducting plate with a dextral sense of shear, and shear strain is accommodated mainly in the regions above the plate interface. In such regions, serpentine [100] axes are oriented parallel to the dip of the slab (shear direction); poles to the (001) plane are oriented normal to the shear plane, and [010] axes are oriented parallel to the trench axis (Fig. 4). Consequently, a shear wave polarized normal to the trench is much slower than a trench-parallel wave, resulting in polarization of the fast shear wave parallel to the trench axis (trench-parallel anisotropy).

The magnitude of detectable shear-wave splitting in the horizontal plane is approximately half of the maximum shear-wave anisotropy generated across the serpentine basal plane and the other direction, because the slowest [001] axis is tilted 45° from the horizontal plane. This is a minimum estimate of the seismic anisotropy, because the serpentine [001] axes may be slightly tilted towards the continent side of the shear zone, and a greater contribution of the [001] axes within the horizontal plane results in the slower seismic velocity being oriented normal to the trench. Moreover, natural strain can be much larger than that in laboratory experiments, suggesting in turn that the strength of anisotropy might also be much greater in natural systems.

From the above results, we conclude that alignment of serpentine in the mantle wedge results in a strong trench-parallel anisotropy ($A_V \approx 32\%$), whereas olivine crystal-preferred orientation produces a much weaker anisotropy ($A_V \approx 2\text{--}5\%$)⁶, resulting in turn in shear-wave splitting with a relatively short delay time. The strong trench-parallel anisotropy observed in other subduction systems, including the Tonga–Kermadec arc⁵, may also result from the serpentine crystal-preferred orientation in association with a steeply dipping slab.

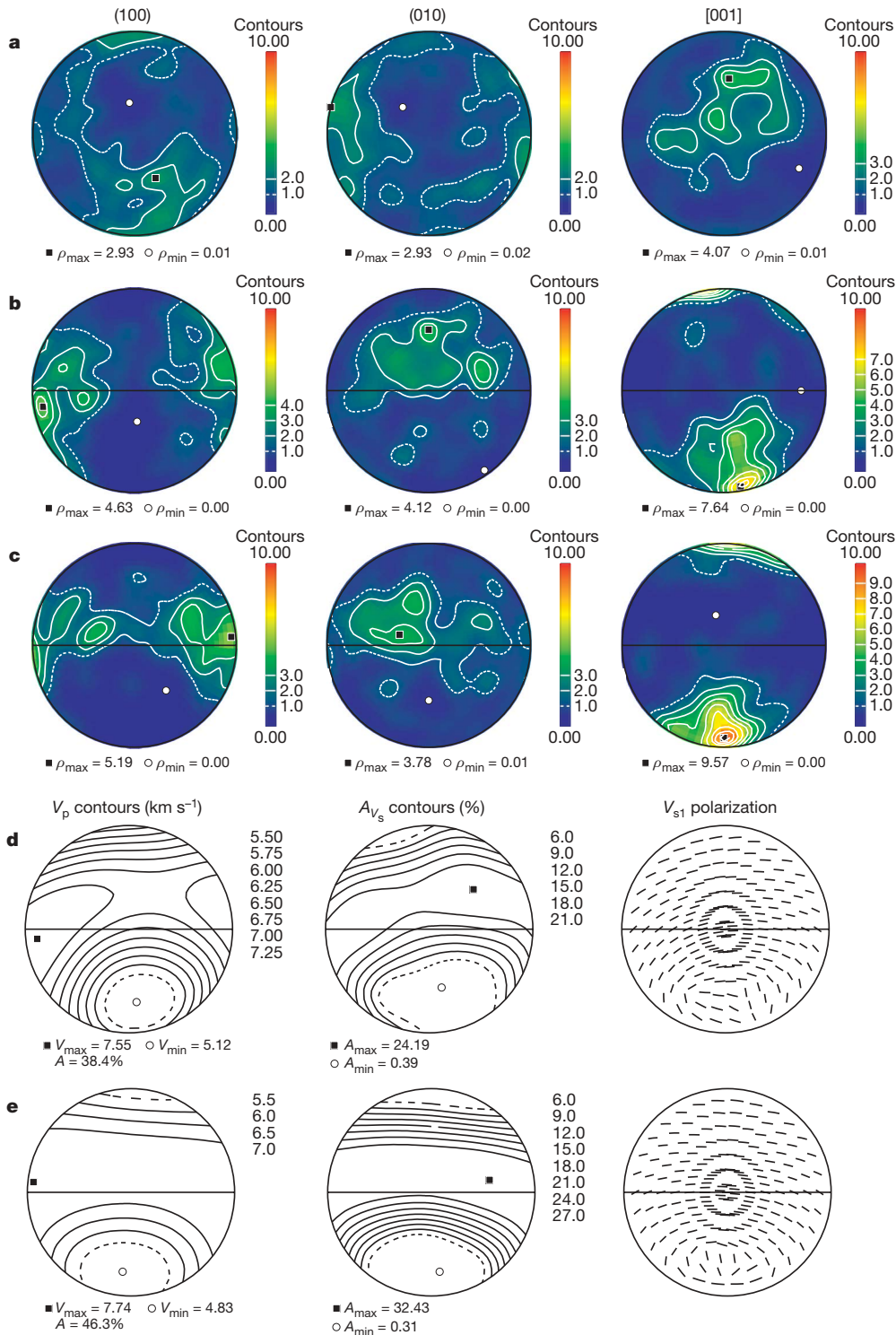


Figure 3 | Pole figures and seismic anisotropy of experimentally deformed serpentine. **a**, Pole figure of crystallographic orientation of the starting material. **b**, Sample deformed at 400 °C and $\gamma = 0.8$. **c**, Sample deformed at 300 °C and $\gamma = 2.0$. For the experimentally deformed samples, the east–west direction corresponds to the shear direction (sense of shear is dextral), and north–south is normal to the shear plane. Both experimental runs are characterized by a strong concentration of [001] axes normal to the shear plane and a relatively weak concentration of [100] axes subparallel to the shear direction. The other two runs show similar patterns, but with slightly weaker fabrics. The pole figures are equal-area lower-hemisphere projections with a Gaussian half-width angle of 8.5°, and the contours represent the density of data points (multiples of uniform density, shading is linear) of 150–200 grains for each sample. Maximum and minimum density (ρ) values are shown. **d**, Seismic anisotropy corresponding to the sample deformed at 400 °C and $\gamma = 0.8$. **e**, Seismic anisotropy corresponding to the sample deformed at 300 °C and $\gamma = 2.0$. Seismic anisotropy was calculated using the elastic constants for lizardite¹¹. Maximum and minimum velocity (V) and anisotropy (A) values are shown. V_p , compressional-wave velocity; V_{s1} , fast shear-wave velocity.

Several models have been proposed to explain the origin of trench-parallel anisotropy in subduction zones, including slab rollback¹⁹, three-dimensional corner flow²⁰, convection driven by crustal delamination²¹, and deformation under water-rich conditions²²; however, in these previous models, seismic anisotropy is mainly attributed to the crystal-preferred orientation of olivine. In contrast, we proposed that strong trench-parallel anisotropy with delay time of 1–2 s can result from the crystal-preferred orientation of serpentine in the hydrated mantle wedge. Even if trench-parallel flow is dominant in the forearc mantle^{5,20}, the basal plane of serpentine becomes aligned subparallel to the flow direction, thereby producing a strong trench-parallel anisotropy. In contrast, for mantle corner flow with a shallowly dipping slab,

seismic anisotropy becomes weak as a result of a smaller contribution from the [001] axes within the horizontal plane.

Trench-parallel anisotropy has recently been ascribed to the occurrence of serpentine-filled faults in the subducting slab²³. The poor vertical resolution of seismic anisotropy data⁸ means that we cannot test this hypothesis; however, our model predicts that the anisotropy is mainly generated by flow in the mantle wedge, consistent with measurements of local S-waves sampled mostly through the mantle wedge⁴.

The spatially variable nature of trench-parallel anisotropy beneath northeast Japan (delay time of 0.1–0.2 s)⁹ and the Ryukyu arc (delay time of 1–2 s)⁴ may reflect the heterogeneous distribution of serpentine.

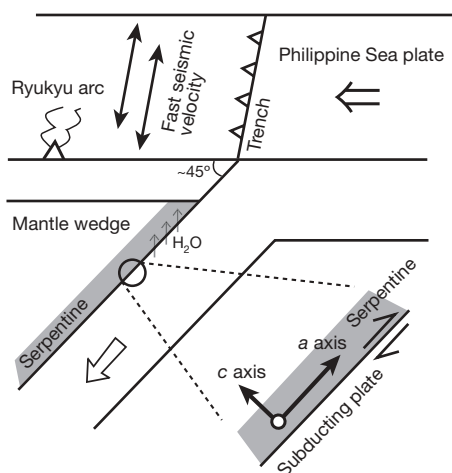


Figure 4 | Schematic cross-section through the Ryukyu arc. Serpentine forms in the mantle wedge owing to the infiltration of water expelled from the subducting plate, and deformation is concentrated in a relatively thin layer above the subducting plate. In this layer, the seismically slowest axis in serpentine (*c* axis) tends to align normal to the plate interface, resulting in a strong trench-parallel anisotropy in the case of steeply subducting slab, such as that beneath the Ryukyu arc.

Serpentinization of the mantle wedge is primarily controlled by the subduction geotherm²⁴; in a relatively warm subduction zone, such as the Ryukyu arc, large amounts of water are released at shallow depths (~40 km) via dehydration reactions that occur in the subducting plate, with the fluids migrating into the overlying mantle and causing extensive serpentinization. In contrast, in cool subduction zones, most dehydration reactions are limited to deeper levels (>100 km) and serpentinization is not expected to occur in the shallow parts of the mantle wedge, such as that beneath northeast Japan.

Seismic tomography reveals an extensive low-velocity anomaly in the mantle wedge beneath the Ryukyu arc⁷, suggesting the occurrence of serpentine, whereas such an anomaly is rarely observed in the forearc mantle beneath northeast Japan²⁵. Our model does not resolve the anisotropic source of shear-wave splitting with a short delay time, but our results indicate that a strong trench-parallel anisotropy (delay time of 1–2 s) can be produced by the crystal-preferred orientation of serpentine in a hydrated mantle wedge.

METHODS SUMMARY

Deformation experiments of serpentine were carried out using a solid-medium apparatus at pressure of 1 GPa and temperature of 300–400 °C. We used natural antigorite as a starting material, which shows a nearly random crystallographic orientation. The thin sliced sample (~400 μm thick) was placed into alumina piston cutting at 45° from the maximum compression direction, and the specimen was deformed at constant rates ranging from 1.3×10^{-4} to $2.5 \times 10^{-4} \text{ s}^{-1}$. The samples were quenched rapidly after each experiment by switching off the thermo-controller to freeze the deformation microstructure, and the recovered sample was cut along the compressional direction. Crystal orientation of serpentine was analysed by EBSD in a scanning electron microscope, and the seismic anisotropy corresponding to the deformed serpentine aggregates was calculated using the software written by D. Mainprice.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 May; accepted 14 September 2009.

1. Nicolas, A. & Christensen, N. I. in *Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System* (eds Fuchs, K. & Froideveaux, C.) 407–433 (American Geophysical Union, 1987).

2. Long, M. D. & Silver, P. G. The subduction zone flow field from seismic anisotropy: a global view. *Science* **319**, 315–318 (2008).
3. Zhang, S. & Karato, S. Lattice preferred orientation of olivine aggregates deformed in simple shear. *Nature* **375**, 774–777 (1995).
4. Long, M. D. & van der Hilst, R. D. Shear wave splitting from local events beneath the Ryukyu arc: trench-parallel anisotropy in the mantle wedge. *Phys. Earth Planet. Inter.* **155**, 300–312 (2006).
5. Smith, G. P. *et al.* A complex pattern of mantle flow in the Lau backarc. *Science* **292**, 713–716 (2001).
6. Katayama, I. & Karato, S. Effect of temperature on the B- to C-type olivine fabric transition and implication for flow pattern in subduction zones. *Phys. Earth Planet. Inter.* **157**, 33–45 (2006).
7. Wang, Z., Huang, R., Huang, J. & He, Z. P-wave velocity and gradient images beneath the Okinawa trough. *Tectonophysics* **455**, 1–13 (2008).
8. Wiens, D. A., Conder, J. A. & Faul, U. H. The seismic structure and dynamics of the mantle wedge. *Annu. Rev. Earth Planet. Sci.* **36**, 421–455 (2008).
9. Nakajima, J. & Hasegawa, A. Shear-wave polarization anisotropy and subduction-induced flow in the mantle wedge of northeastern Japan. *Earth Planet. Sci. Lett.* **225**, 365–377 (2004).
10. Katayama, I. Thin anisotropic layer in the mantle wedge beneath northeast Japan. *Geology* **37**, 211–214 (2009).
11. Auzende, A. L., Pellenq, R. J. M., Devouard, B., Baronnet, A. & Grauby, O. Atomistic calculations of structural and elastic properties of serpentine minerals: the case of lizardite. *Phys. Chem. Miner.* **33**, 266–275 (2006).
12. Abramson, E. H., Brown, J. M., Slutsky, L. J. & Zaug, J. The elastic constants of San Carlos olivine to 17 GPa. *J. Geophys. Res.* **102**, 12253–12263 (1997).
13. Hyndman, R. D. & Peacock, S. M. Serpentinization of the forearc mantle. *Earth Planet. Sci. Lett.* **212**, 417–432 (2003).
14. Kneller, E. A., Long, M. D. & van Keken, P. E. Olivine fabric transitions and shear wave anisotropy in the Ryukyu subduction system. *Earth Planet. Sci. Lett.* **268**, 268–282 (2008).
15. Hilaret, N. *et al.* High-pressure creep of serpentine, interseismic deformation, and initiation of subduction. *Science* **318**, 1910–1913 (2007).
16. Kern, H., Liu, B. & Popp, T. Relation between anisotropy of P and S wave velocities and anisotropy of attenuation in serpentinite and amphibolite. *J. Geophys. Res.* **102**, 3051–3065 (1997).
17. Watanabe, T., Kasami, H. & Ohshima, S. Compressional and shear wave velocities of serpentinized peridotites up to 200 MPa. *Earth Planets Space* **59**, 233–244 (2007).
18. Li, C., van der Hilst, R. D. & Toksoz, M. N. Constraining P-wave velocity variations in the upper mantle beneath Southeast Asia. *Phys. Earth Planet. Inter.* **154**, 180–195 (2006).
19. Buttle, J. & Olson, P. A. A laboratory model of subduction zone anisotropy. *Earth Planet. Sci. Lett.* **164**, 245–262 (1998).
20. Kneller, E. A. & van Keken, P. Trench-parallel flow and seismic anisotropy in the Mariana and Andean subduction systems. *Nature* **450**, 1222–1225 (2007).
21. Behn, M. D., Hirth, G. & Kelemen, P. Trench-parallel anisotropy produced by foundering of arc lower crust. *Science* **317**, 108–111 (2007).
22. Jung, H. & Karato, S. Water-induced fabric transitions in olivine. *Science* **293**, 1460–1463 (2001).
23. Faccenda, M., Burlini, L., Gerya, T. & Mainprice, D. Fault-induced seismic anisotropy by hydration in subducting oceanic plates. *Nature* **455**, 1097–1101 (2008).
24. Hacker, B. R., Abers, G. A. & Peacock, S. M. Subduction factory. 1. Theoretical mineralogy, densities, seismic wave speeds, and H₂O contents. *J. Geophys. Res.* **108**, doi:10.1029/2001JB001127 (2003).
25. Nakajima, J., Matsuzawa, T., Hasegawa, A. & Zhao, D. Three-dimensional structure of Vp, Vs and Vp/Vs beneath northeast Japan: implications for arc magmatism and fluids. *J. Geophys. Res.* **106**, 21843–21857 (2001).
26. Nakajima, J., Shimizu, J., Hori, S. & Hasegawa, A. Shear-wave splitting beneath the southwestern Kurile arc and northeastern Japan arc. *Geophys. Res. Lett.* **33**, L05305 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Watanabe and S. Karato for comments and discussions. We also thank D. Mainprice and A. Tommasi for providing the crystallographic data for serpentine. This study was supported by the Japan Society for the Promotion of Science (JSPS).

Author Contributions I.K. and K.-i.H. planned the project and performed the experiments. K.M. has responsibility for the EBSD analysis, and J.-i.A. for the experimental procedures. All authors discussed the results and implications.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.K. (katayama@hiroshima-u.ac.jp).

METHODS

Experiments were conducted in a simple shear geometry using a solid-medium apparatus at Hiroshima University, Japan. The starting materials of serpentinite were collected from the Sambagawa metamorphic belt, which consist mainly of antigorite (95 vol.%) and magnetite (<5 vol.%) with a nearly random orientation. The specimen used for deformation experiments were prepared by first drilling a cylinder of diameter 4.0 mm and then cutting off a ~400- μ m-thick slice at an angle of 45° to the long axis. The sample was then cut into half normal to the plane and a nickel strain-marker was installed. The serpentine sample was sandwiched between the alumina pistons that cut at 45° from the maximum compression direction, and surrounded by a nickel jacket (Supplementary Figure).

The pressure was first raised to 400 MPa, and the temperature was increased to the desired value (300–400 °C), and then the pressure was increased again to

1,000 MPa. Temperature was monitored by two Pt/Rh thermocouples, and the variation reading from two thermocouples was usually within ~10–20 °C. After temperature and pressure reached the desired levels, a piston was advanced at a constant rate using a servo-motor²⁷. The samples were quenched rapidly after experiments by switching off the thermo-controller and then pressure was reduced. Shear strain was measured by the rotation of the nickel strain-marker in the recovered sample. We also measured the sample thickness before and after each experiment, and found that the sample thickness changed about 10% during experiments. The strain rate was calculated from the obtained strain and the duration of the deformation. The experimental conditions are summarized in the Supplementary Table.

27. Ando, J., Takeshita, T., Matsubara, K. & Hayasaka, Y. Evaluation of fundamental performance of a modified Griggs type apparatus installed at Hiroshima University. *Jap. J. Struct. Geol.* **49**, 27–39 (2006).

LETTERS

Convergent evolution of anthropoid-like adaptations in Eocene adapiform primates

Erik R. Seiffert¹, Jonathan M. G. Perry², Elwyn L. Simons³ & Doug M. Boyer⁴

Adapiform or 'adapoid' primates first appear in the fossil record in the earliest Eocene epoch (~55 million years (Myr) ago), and were common components of Palaeogene primate communities in Europe, Asia and North America¹. Adapiforms are commonly referred to as the 'lemur-like' primates of the Eocene epoch, and recent phylogenetic analyses have placed adapiforms as stem members of Strepsirrhini^{2–4}, a primate suborder whose crown clade includes lemurs, lorises and galagos. An alternative view is that adapiforms are stem anthropoids⁵. This debate has recently been rekindled by the description of a largely complete skeleton of the adapiform *Darwinius*⁶, from the middle Eocene of Europe, which has been widely publicised as an important 'link' in the early evolution of Anthropoidea⁷. Here we describe the complete dentition and jaw of a large-bodied adapiform (*Afradapis* gen. nov.) from the earliest late Eocene of Egypt (~37 Myr ago) that exhibits a striking series of derived dental and gnathic features that also occur in younger anthropoid primates—notably the earliest catarrhine ancestors of Old World monkeys and apes. Phylogenetic analysis of 360 morphological features scored across 117 living and extinct primates (including all candidate stem anthropoids) does not place adapiforms as haplorhines (that is, members of a *Tarsius*–Anthropoidea clade) or as stem anthropoids, but rather as sister taxa of crown Strepsirrhini; *Afradapis* and *Darwinius* are placed in a geographically widespread clade of caenopithecine adapiforms that left no known descendants. The specialized morphological features that these adapiforms share with anthropoids are therefore most parsimoniously interpreted as evolutionary convergences. As the largest non-anthropoid primate ever documented in Afro-Arabia, *Afradapis* nevertheless provides surprising new evidence for prosimian diversity in the Eocene of Africa, and raises the possibility that ecological competition between adapiforms and higher primates might have played an important role during the early evolution of stem and crown Anthropoidea in Afro-Arabia.

Birket Qarun Locality 2 (BQ-2) is a near-coastal fluvial freshwater deposit exposed in the Fayum Depression of northern Egypt that has produced a diverse mammalian fauna that includes primitive anthropoid⁴ and crown strepsirrhine⁸ primates. An age estimate of ~37 Myr ago (that is, earliest late Eocene or earliest Priabonian) for the locality is supported by palaeomagnetic reversal stratigraphy^{4,9} as well as sequence stratigraphic and invertebrate biostratigraphic evidence from adjacent rocks^{10,11}. The new genus and species of adapiform described here is one of the more common mammals at BQ-2, and is, by far, the largest primate from the locality.

Primates Linnaeus, 1758

Adapiformes Hoffstetter, 1977

Caenopithecinae Godinot, 1998

Afradapis longicristatus gen. et sp. nov.

Etymology. Generic name derived from *Afra*, Latin, Africa, and *adapis*, in reference to *Adapis* Cuvier, 1821. Specific epithet is from *longi*-, Latin, long, and *cristatus*, Latin, crested.

Holotype. CGM 83690, partial left mandible preserving P₄–M₃ and masseteric fossa (Fig. 1r).

Locality and horizon. Early Priabonian (early late Eocene) Birket Qarun Locality 2 (BQ-2), Umm Rigl Member, Birket Qarun Formation, Fayum Depression, northern Egypt.

Diagnosis. Large adapiform (mean body mass estimates of 2,187 g ('prosimian' regression) and 3,262 g ('all primate' regression) based on M₁ area¹²; 3,283 g based on M₂ length¹³) that differs from other caenopithecine adapiforms in exhibiting the following combination of characters: absence of P₂; tall and trenchant upper and lower third premolars; mesoconid variably present on lower molars, particularly on M₁; no metastylids on M_{1–3}; mesostyles, large hypocones and prehypocristae present on all upper molars; mandible with deep and short corpus, extensive masseteric fossa, ossified symphysis with transverse torus, short condylar neck and low condyle relative to the mandibular tooth row.

Description. Mandibular fragments collectively preserve the canine and postcanine tooth row and most of the corpus anterior to the M₃, and definitively demonstrate that P₂ was absent (Figs 1q and 2). The corpus is very deep, the symphysis is fully ossified in DPC 21370J and DPC 24085A, and both specimens bear transverse tori superior and inferior to the genial tubercles (Fig. 1s–u). The lingual symphyseal shelf is long, extending back to the level of P₄, and is oriented at roughly 45 degrees to the occlusal plane.

The lower incisor crowns of *Afradapis* are spatulate as in other known adapiforms, and bear lingual keels and weak lingual cingulids (Fig. 1m, n). The I₁ has a flat and sharp occlusal edge. The I₂ occlusal edge is in line with that of I₁ mesially, but then curves steeply distally. The alveoli for I_{1–2} are only slightly procumbent. The upper incisors (Fig. 1k, l) are asymmetric and have occlusal surfaces that are rounded to either side of a lingual keel; their edges probably did not form a continuous blade. The upper central incisor crowns have well-developed mesial processes, but none of the isolated upper incisors has an interproximal wear facet, indicating that there was probably no midline contact and at least a small interincisal gap. The lower canine (Fig. 1v) is robust, protrudes high above the occlusal plane and bears a distal heel that is often worn from contact with the anterior edge of the upper canine. Many of the upper lateral incisors have a strong wear facet on their distal surfaces from contact with the lower canine. The upper canine has sharp mesial and distal edges, a weak lingual cingulum and a deep mesial groove (Fig. 1o). There is no clear evidence for sexual dimorphism in canine size.

The P₃ is particularly notable for being longer and taller than P₄ (Fig. 1j, q), and for having a long honing facet that develops from wear against the distal face of the upper canine. The P₃ roots are implanted

¹Department of Anatomical Sciences, Stony Brook University, Stony Brook, New York 11794-8081, USA. ²Department of Anatomy, Midwestern University, Downers Grove, Illinois 60515, USA. ³Division of Fossil Primates, Duke Lemur Center, 1013 Broad Street, Durham, North Carolina 27705, USA. ⁴Department of Ecology & Evolution, Stony Brook University, Stony Brook, New York 11794-5245, USA.

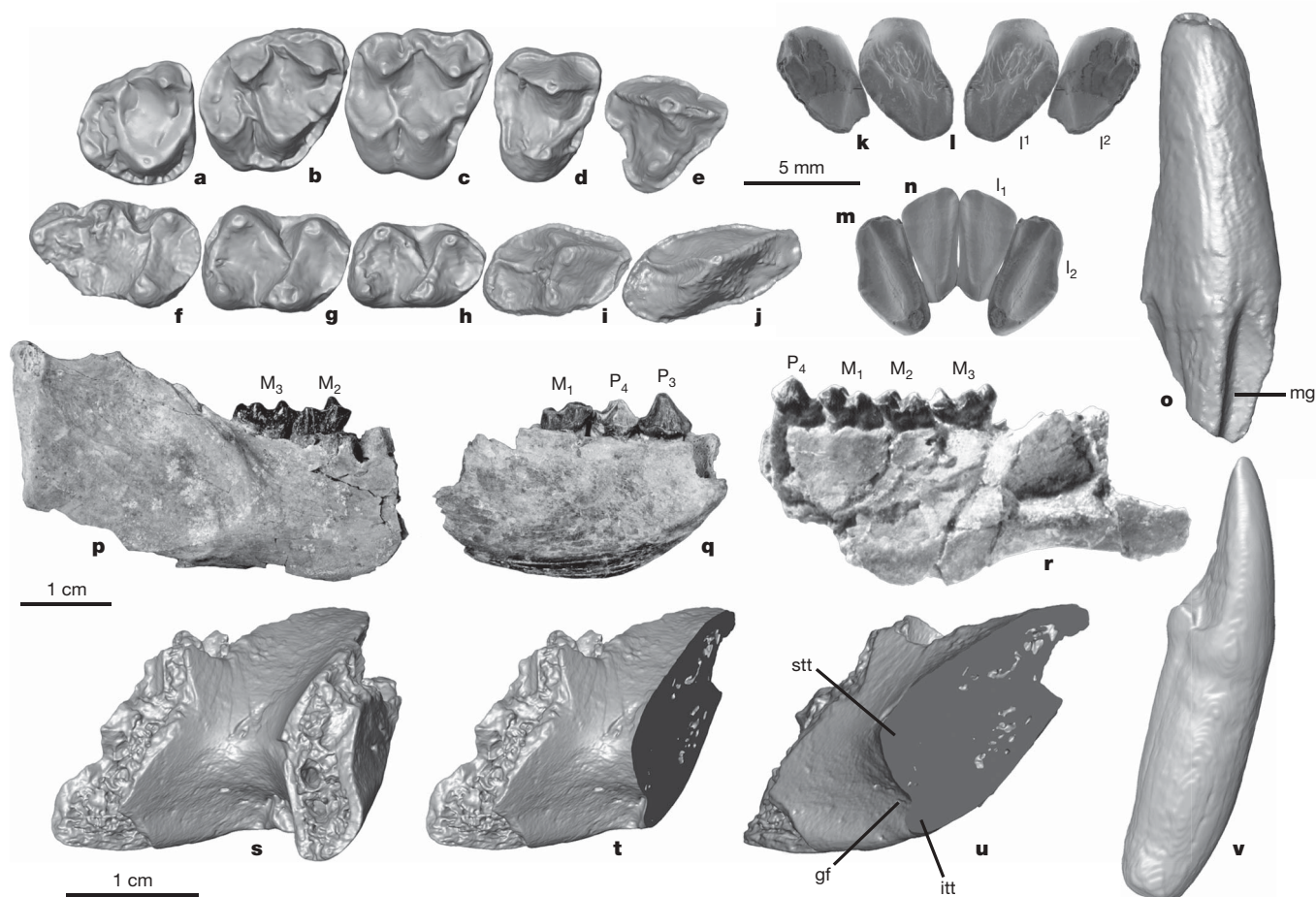


Figure 1 | Dental and mandibular remains of *Afradapis longicristatus* gen. et sp. nov. a, M^3 (DPC 21458D); b, M^2 (DPC 21370B); c, M^1 (DPC 21366A); d, P^4 (DPC 21370C); e, P^3 (DPC 21500G); f, M^3 (DPC 21366F, reversed); g, M^2 (DPC 21366E); h, M^1 (DPC 21456C); i, P^4 (DPC 21548I); j, P^3 (DPC 21500I, reversed); k, I^2 (DPC 22440D); incisors are mirror-imaged to reconstruct incisor battery; l, I^1 (DPC 22442K); m, I_2 (DPC 21456G); n, I_1 (DPC 21639E); o, upper canine (DPC 21578J) in oblique buccal view; p, mandible with M_{2-3} (DPC 23275); q, mandible with P_3 – M_1 and lower canine alveolus (DPC

21458F); r, holotype lower mandible with P_4 – M_3 (CGM 83690); s–u, micro-CT reconstruction of edentulous mandible preserving fused symphysis and transverse tori (DPC 21370J) in oblique posterior (s, t) and medial (u) views (specimen is digitally sliced along the midline of the symphysis in t and u); v, lower canine (DPC 21365E) in buccal view. Five millimetre scale bar is for a–o and v; upper 1 cm scale bar is for p–r; lower 1 cm scale bar is for s–u. gf, Genial fossa; i, incisor; itt, inferior transverse torus; m, molar; mg, mesial groove; p, premolar; stt, superior transverse torus.

obliquely with respect to those of P_4 , and the long axes of the primary crests are strongly oblique with respect to the long axis of the tooth row. Among primates, this derived complex of features, along with the loss of P_2 , is otherwise seen only in living and extinct catarrhine anthropoids, and is first documented among stem catarrhines three

million years later in the Fayum area¹⁴. Other caenopithecines retain a small, single-rooted P_2 (refs 6, 15–18). The P_4 (Fig. 1i, q, r) has a prominent, distolingually placed metaconid and a large talonid with a distinct cristid obliqua. Only P^3 and P^4 of *Afradapis* have been recovered as isolated teeth, consistent with P^2 being absent. The P^3 (Fig. 1e) has a tall, blade-like paracone and a very small protocone. The P^4 (Fig. 1d) is slightly waisted and has a relatively well-developed protocone.

The upper and lower molars of *Afradapis* are superficially similar to those of some highly folivorous extant primates, such as those of platyrrhine howler monkeys (*Alouatta*) and lemuriform indriids such as the sifaka (*Propithecus*) and indri (*Indri*). *Afradapis*' elaborate shearing crests indicate that the species was likely to have been highly folivorous as well. The lower molars of *Afradapis* (Fig. 1f–h) lack paraconids, but otherwise have high cusps and long crests, and are often densely crenulated. Lower molar length increases, and the relative length of the trigonid basins decreases, distally along the toothrow. The hypoconulid is very small or absent on M_{1-2} , but the M_3 has a large hypoconulid lobe. The upper molars (Fig. 1a–c) have moderately tall cusps and long, sharp crests. M^{1-2} have well-developed mesostyles, distinct buccal cingula, large hypocones (derived from the posterior cingulum), and trenchant postprotocristae and prehypocristae. A large paraconule is present along the preprotocrista, but there is no metaconule. Upper molars decrease in size, and have more rounded distal margins, moving

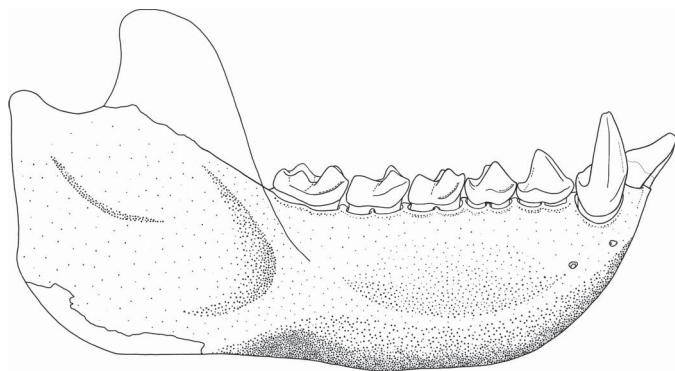


Figure 2 | Reconstruction of the mandible and lower dentition of *Afradapis longicristatus*, gen. et sp. nov. Reconstruction is based on combined information from specimens CGM 83690, DPC 23275, DPC 21458F and DPC 21370J.

distally along the toothrow. The M^3 bears an expanded talon and a strongly convex distal margin, and has relatively small metacone, hypocone and paraconule cusps.

It has long been known that some adapiform lineages evolved derived morphological features that are also seen in living and extinct anthropoids (for example, fused mandibular symphyses, upper canines with mesial grooves, enlarged and spatulate upper and lower incisors, short and tall rostra)¹⁶. The phylogenetic significance of these features has been a source of ongoing debate for decades^{3,5,19,20}. Of all known fossil prosimians (including *Darwinius*), *Afradapis* provides perhaps the most detailed examples of derived anthropoid-like adaptations in its dental and mandibular morphology. As is the case for many of the morphological features that some have argued link adapiforms to anthropoids, however, the anthropoid-like features of *Afradapis* (fused mandibular symphysis with transverse torus, deep mandibular corpus, deep masseteric fossa, large upper molar hypocones, absence of $P^2/2$ and presence of an enlarged P_3 with a honing facet for the upper canine) are not present in the most primitive undoubted fossil anthropoids, such as *Biretia*⁴ and *Proteopithecus*²¹, indicating that the features are likely to have been acquired through convergent evolution.

Despite the key role that adapiforms have had in debates surrounding anthropoid and strepsirrhine origins, the group has been poorly sampled in recent analyses of primate interrelationships. We expanded on previous analyses^{4,22} by scoring 117 living and extinct primates (including 24 adapiforms, eight of which were not sampled previously) for 360 morphological features and analysed the resulting data set using the heuristic search algorithm in PAUP* 4.0b10 (ref. 23). Parsimony analysis places *Afradapis* and its caenopithecine relatives, including *Darwinius*, within a large clade of adapiform primates that includes European adapines and 'cercamoniines' and North American notharctines. This group, along with Asian sivaladapids and asiadapines²⁴, is recovered as one of several sister clades of crown Strepsirrhini (Fig. 3; for methods see Methods; for character support, see Supplementary Information). Interestingly, *Afradapis* and the younger (~34 Myr ago) African caenopithecine *Aframoni* are placed as consecutive sister taxa of European middle Eocene *Caenopithecus*, indicating at least one dispersal across the Tethys Sea. Given that there is no other compelling evidence for mammalian faunal exchange between Europe and Afro-Arabia during the middle Eocene^{25,26}, it is possible that the few derived dental features that *Afradapis* and *Caenopithecus* share to the exclusion of *Aframoni* evolved independently on the two landmasses from more generalized caenopithecine ancestors. More complete fossil evidence from the middle Eocene of Europe and Afro-Arabia will be required to provide a robust test of these alternative hypotheses.

The *Darwinius* holotype presents unique problems for phylogenetic analysis because the specimen is a juvenile, most of its bones are crushed or distorted, and many of the characters that have had a key role in primate phylogenetics (that is, features of the dentition, auditory region and ankle) cannot be evaluated. For instance, in our opinion the orientation of the fibular facet on the astragalus cannot be determined with confidence (*contra* ref. 6) because the astragalus is damaged, the orientation of the trochlear surface cannot be determined, and the fibular facet is almost entirely obscured by the fibular malleolus (see Fig. 11 in ref. 6). However, the topology presented in Fig. 3 does not change if *Darwinius* is scored as having a straight-sided fibular facet like haplorhines (and unlike all other adapiforms for which this region is known). Constraining caenopithecine adapiforms to be placed closer to crown Anthropoidea than two other groups (Eosimiidae and Amphipithecidae from the Paleogene of Asia^{27–29}) that emerge as stem anthropoids in the unconstrained tree adds 12 steps to tree length (see Supplementary Information). Even in this much longer tree, *Darwinius* and other caenopithecines are most parsimoniously placed in a clade with more primitive adapiforms such as adapines, asiadapines and 'cercamoniines', and not as intermediate forms that might link these taxa to undoubted anthropoids.

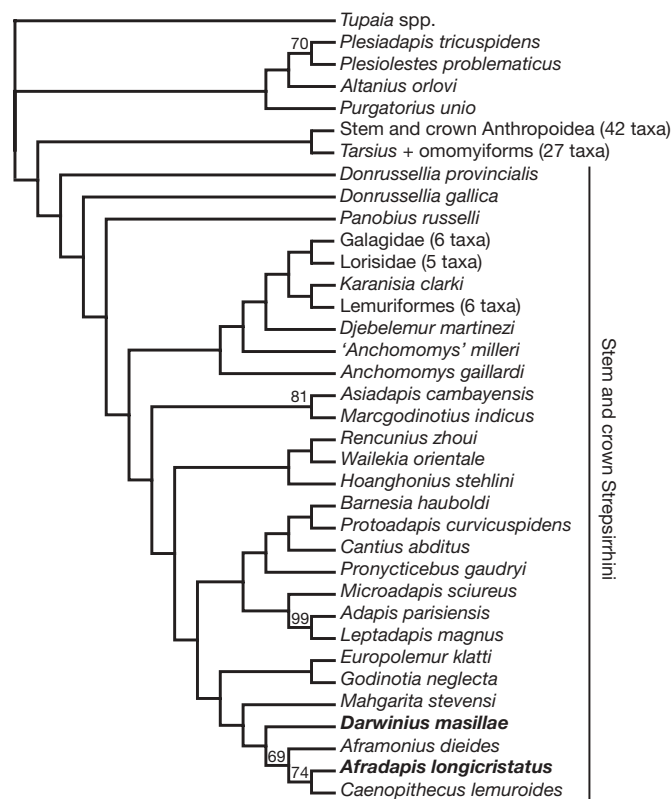


Figure 3 | Phylogenetic position of the adapiforms *Afradapis* and *Darwinius* within Primates. Single tree recovered from parsimony analysis of a morphological character matrix containing 360 characters scored across 117 taxa. Tree length = 2,265.671; consistency index excluding uninformative characters = 0.1891; retention index = 0.5848; rescaled consistency index = 0.1120. Values above branches are bootstrap support, based on 1,000 pseudoreplicates. See Supplementary Information for full tree, matrix, methods and tests of alternative topologies.

The adapiform radiation produced an impressive range of morphologically diverse taxa whose evolutionary history is well documented in North America, Europe and, to a lesser extent, Asia. Adapiforms were not definitively known from Africa until the last decade, but it now seems that members of the group were not only present in Afro-Arabia, but were probably common and quite successful living alongside, and competing with, early anthropoid primates in the middle and late Eocene. Our analysis shows that adapiforms are unlikely to have given rise to Anthropoidea, but the former were evidently the first primates to invade and occupy anthropoid-like feeding niches in Africa. As early ecological competitors of stem and crown anthropoids, adapiforms may still have an important role for understanding the selection pressures that drove the origin and early evolution of Anthropoidea.

METHODS SUMMARY

The morphological character matrix analysed in this study is a modified version of the matrix compiled and analysed in ref. 4. The results presented in Fig. 3 are based on parsimony analysis of the matrix with some multistate characters treated as ordered, and with ordered characters scaled, so that any single change within a character could only contribute a maximum of one step to tree length. For many characters, taxa exhibiting intraspecific variation were assigned an intermediate character state, rather than a traditional polymorphic (for example, '0/1') coding. Some characters documenting differences in premolar number were constrained by step matrices that did not allow premolar teeth to be re-acquired following an earlier loss. A molecular scaffold was implemented based on the results of molecular studies, particularly analysis of short interspersed elements (SINEs), that strongly support the monophyly of Malagasy lemuriforms, and, within Lorisiformes, a monophyletic Lorisidae and an *Arctocebus-Perodicticus* clade³⁰. The matrix was analysed with random addition sequence and TBR branch swapping across 2,000 replicates. Tree lengths were recalculated

after networks were re-rooted with *Tupaia* spp. Bootstrap support derives from 1,000 pseudoreplicates.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 July; accepted 18 August 2009.

- Godinot, M. A summary of adapiform systematics and phylogeny. *Folia Primatol. (Basel)* **69** (suppl. 1), 218–249 (1998).
- Marivaux, L. *et al.* Anthropoid primates from the Oligocene of Pakistan (Bugti Hills): data on early anthropoid evolution and biogeography. *Proc. Natl Acad. Sci. USA* **102**, 8436–8441 (2005).
- Ross, C., Williams, B. & Kay, R. F. Phylogenetic analysis of anthropoid relationships. *J. Hum. Evol.* **35**, 221–306 (1998).
- Seiffert, E. R. *et al.* Basal anthropoids from Egypt and the antiquity of Africa's higher primate radiation. *Science* **310**, 300–304 (2005).
- Rasmussen, D. T. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 335–360 (Plenum, 1994).
- Franzen, J. L. *et al.* Complete primate skeleton from the middle Eocene of Messel in Germany: morphology and paleobiology. *PLoS ONE* **4**, e5723 (2009).
- Tudge, C. *The Link: Uncovering Our Earliest Ancestor* (Little, Brown & Company, 2009).
- Seiffert, E. R., Simons, E. L. & Attia, Y. Fossil evidence for an ancient divergence of lorises and galagos. *Nature* **422**, 421–424 (2003).
- Seiffert, E. R. Revised age estimates for the later Paleogene mammal faunas of Egypt and Oman. *Proc. Natl Acad. Sci. USA* **103**, 5000–5005 (2006).
- Gingerich, P. D. Marine mammals (Cetacea and Sirenia) from the Eocene of Gebel Mokattam and Fayum, Egypt: stratigraphy, age, and paleoenvironments. *Univ. Mich. Pap. Paleont.* **30**, 1–84 (1992).
- Seiffert, E. R., Bown, T. M., Clyde, W. C. & Simons, E. L. in *Elwyn L. Simons: A Search for Origins* (eds Fleagle, J. G. & Gilbert, C. C.) 71–86 (Springer, 2008).
- Conroy, G. C. Problems of body-weight estimation in fossil primates. *Int. J. Primatol.* **8**, 115–137 (1987).
- Kay, R. F. & Simons, E. L. The ecology of Oligocene African Anthropoidea. *Int. J. Primatol.* **1**, 21–37 (1980).
- Simons, E. L. Discovery of the oldest known anthropoidean skull from the Paleogene of Egypt. *Science* **247**, 1567–1569 (1990).
- Simons, E. L., Rasmussen, D. T. & Gingerich, P. D. New cercamoniine adapid from Fayum, Egypt. *J. Hum. Evol.* **29**, 577–589 (1995).
- Rasmussen, D. T. The phylogenetic position of *Mahgarita stevensi*: protoanthropoid or lemuroid? *Int. J. Primatol.* **11**, 439–469 (1990).
- Wilson, J. A. & Szalay, F. S. New adapid primate of European affinities from Texas. *Folia Primatol. (Basel)* **25**, 294–312 (1976).
- Stehlin, H. G. Die Säugetiere des schweizerischen Eocaens. *Abh. Schweiz. Paläont. Ges.* **41**, 1299–1552 (1916).
- Ross, C. F. in *Anthropoid Origins* (eds Fleagle, J. G. & Kay, R. F.) 469–548 (Plenum, 1994).
- Simons, E. L. & Rasmussen, D. T. Skull of *Catopithecus browni*, an early Tertiary catarrhine. *Am. J. Phys. Anthropol.* **100**, 261–292 (1996).
- Miller, E. R. & Simons, E. L. Dentition of *Proteopithecus sylviae*, an archaic anthropoid from the Fayum, Egypt. *Proc. Natl Acad. Sci. USA* **94**, 13760–13764 (1997).
- Seiffert, E. R., Simons, E. L., Ryan, T. M. & Attia, Y. Additional remains of *Wadilemur elegans*, a primitive stem galagid from the late Eocene of Egypt. *Proc. Natl Acad. Sci. USA* **102**, 11396–11401 (2005).
- Swofford, D. L. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4 (Sinauer Associates, 1998).
- Rose, K. D. *et al.* Early Eocene Primates from Gujarat, India. *J. Hum. Evol.* **56**, 366–404 (2009).
- Gheerbrant, E. & Rage, J. C. Paleobiogeography of Africa: How distinct from Gondwana and Laurasia? *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **241**, 224–246 (2006).
- Tabuce, R. & Marivaux, L. Mammalian interchanges between Africa and Eurasia: an analysis of temporal constraints on plausible anthropoid dispersals during the Paleogene. *Anthropol. Sci.* **113**, 27–32 (2005).
- Chaimanee, Y., Suteethorn, V., Jaeger, J.-J. & Ducrocq, S. A new late Eocene anthropoid primate from Thailand. *Nature* **385**, 429–431 (1997).
- Jaeger, J.-J. *et al.* A new primate from the Middle Eocene of Myanmar and the Asian early origin of anthropoids. *Science* **286**, 528–530 (1999).
- Beard, K. C., Qi, T., Dawson, M. R., Wang, B. & Li, C. A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. *Nature* **368**, 604–609 (1994).
- Roos, C., Schmitz, J. & Zischler, H. Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl Acad. Sci. USA* **101**, 10650–10654 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. C. Beard, P. Chatrath, B. Engesser, M. Hellmund, J. Galkin, P. D. Gingerich, M. Godinot, G. Gunnell, E. Ladier, B. Marandat, L. Marivaux, K. D. Rose, I. Rutzky and P. Tassy for access to adapiform fossils and casts. F. Ankel-Simons and J. Fleagle provided comments on the manuscript. Collaborative palaeontological research in Egypt is made possible through cooperation with the Egyptian Mineral Resources Authority and the Egyptian Geological Museum. Fieldwork in Egypt is managed by P. Chatrath. This research has been funded by the Research Foundation of SUNY, and grants from the US National Science Foundation and The Leakey Foundation to E.R.S. and E.L.S. This is Duke Lemur Center publication 1155.

Author Contributions E.R.S. and E.L.S. recovered the fossils of *Afradapis* along with other members of their field crews in Egypt; E.R.S. assembled the character matrix and all figures, created Figs 2 and 3, ran the phylogenetic analyses, and compiled the Supplementary Information document; E.R.S. and J.M.G.P. analysed the fossils and wrote the manuscript; D.M.B. and J.M.G.P. CT-scanned fossils and provided three-dimensional reconstructions for Fig. 1. D.M.B. scored plesiadapiforms in the character matrix and helped to write the manuscript. E.L.S. helped to write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.R.S. (erik.seiffert@stonybrook.edu).

METHODS

The character–taxon matrix used in this study is a modified version of the matrix compiled and analysed in ref. 4, expanded to include 360 characters and 117 taxa. Codings for crown primates are based on observations made on original material and casts by E.R.S., in some cases supplemented by codings provided by ref. 3. Characters derive from ref. 31 (abbreviation ‘D’ in matrix), ref. 32, ref. 33 (abbreviation ‘K’ in matrix), ref. 34, ref. 3 (abbreviation ‘R’ in matrix), ref. 35 (abbreviation ‘Y’ in matrix), and ref. 36 (abbreviation ‘YIP’ in matrix). Scorings for *Darwinius masillae* are based on observations made by E.R.S. on high-resolution casts of ‘Plate A’ made available for study by P. Gingerich, supplemented by photographs and micro-CT reconstructions published in ref. 6. Scorings for features preserved on ‘Plate B’ are based on ref. 37. Original adapiform fossil material was examined by E.R.S. at the Geiseltalmuseum (Halle, Germany), the Muséum National d’Histoire Naturelle (Paris, France), the Carnegie Museum of Natural History (Pittsburgh, Pennsylvania, USA), the University of Michigan Museum of Paleontology (Ann Arbor, Michigan, USA), and the Duke Lemur Center Division of Fossil Primates (Durham, North Carolina, USA). J.M.G.P. made observations on adapiform primates housed at the Naturhistorisches Museum Basel (Basel, Switzerland), the Institut des Sciences de l’Évolution, Université Montpellier II (Montpellier, France), the Muséum National d’Histoire Naturelle (Paris, France), and the Muséum d’Histoire Naturelle, Montauban (Montauban, France). Observations on plesiadapiforms by D.M.B. led to a number of corrections to character codings for *Plesiadapis tricuspidens* and *Plesiolestes problematicus* in the matrix published in ref. 4; several other cells previously coded as ‘missing’ for these taxa were also scored based on D.M.B.’s observations on original material and casts. Scorings for the asiadapines *Asiadapis* and *Marcgodinotius* are from ref. 24, supplemented by observations on casts.

The results presented in Fig. 3 of the main text are based on analysis of the matrix with some multistate characters treated as ordered (see character list), and with ordered characters scaled, so that any single change within a character could only contribute a maximum of one step to tree length. For many characters, taxa exhibiting intraspecific variation were assigned an intermediate character state, rather than a traditional polymorphic (for example, ‘0/1’) coding. Some characters documenting differences in premolar number (character 21, ‘P₁ presence’; character 22, ‘P₂ roots/presence’; character 120, ‘P₁ presence’; character 121, P₂ root number/presence’) were constrained by step matrices that did not allow premolar teeth to be reacquired following an earlier loss (see character matrix). Although supernumerary teeth have been acquired in some highly specialized mammalian lineages, such as sirenians³⁸ and whales^{39,40}, we consider reacquisition of previously lost premolars in these primate clades to be highly unlikely, and a reasonable assumption for phylogenetic analysis. A molecular scaffold (*sensu* ref. 41) was implemented based on the results of molecular studies, particularly analysis of SINEs, that strongly support the monophyly of Malagasy lemuriforms, and within Lorisiformes, a monophyletic Lorisidae and an *Arctocebus–Perodicticus* clade³⁰. The molecular scaffold used here is as follows: (*Tupaia* spp., ((((*Arctocebus calabarensis*, *Perodicticus potto*), (*Loris tardigradus*, *Nycticebus coucang*)), (*Galagoides demidoffi*, *Otolemur crassicaudatus*, *Galago moholi*)), ((*Cheirogaleus major*, *Microcebus murinus*), (*Lemur catta*, *Varecia variegata*), *Propithecus* spp., *Lepilemur mustelinus*)), *Tarsius bancanus*, (*Aotus trivirgatus*, *Saimiri sciureus*)). The matrix was analysed using the heuristic search algorithm in PAUP* 4.0b10 (ref. 23), with random addition sequence and TBR branch swapping across 2,000 replicates. In order to replicate the results presented here, the matrix (Part IV of the Supplementary Information) should be saved as a new document with extension ‘*.nex’; the constraint tree (Part V of the Supplementary Information) should be saved as a new document with the same name as the matrix but with extension ‘*.tre’. The matrix should then be loaded within PAUP* (the backbone constraint will load automatically), and analysed using command line text ‘hsearch nreps=2000 addseq=random enforce=yes’. Tree lengths were recalculated after networks were re-rooted with *Tupaia* spp. Bootstrap support derives from 1,000 pseudo-replicates. Templeton and Kishino–Hasegawa tests of alternative hypotheses were calculated in PAUP* 4.0b10. Character 216, ‘morphology of fibular facet’ was not scored for *Darwinius masillae* in the matrices used in searches for alternative topologies.

Scorings for two taxa deserve special mention. The trees in Fig. 3 of the main text and Part II of the Supplementary Information are based on analyses that

include characters from the NMMP 20 partial skeleton⁴² in the *Pondaungia coteri* OTU. There is ongoing debate surrounding the allocation of these postcranial remains⁴³, but they cannot be attributed to any other large-bodied primate taxon known from the Pondaung Formation. Until compelling evidence is provided for the presence of another taxon that could have left these remains, we consider attribution to *Pondaungia* to be the most parsimonious explanation. We did not score the alleged frontal bones that have been attributed to *Amphipithecus*⁴⁴, following the concerns raised in ref. 45. Although the NMMP 20 postcranial bones have been interpreted as providing evidence for the adapiform affinities of amphipithecids⁴⁶, the group is nevertheless still placed along the stem lineage of Anthropeidea in our analysis. Parsimony analysis following exclusion of the character states observable on NMMP 20 places amphipithecids within crown Anthropeidea as the sister group of living and extinct platyrrhine anthropoids, but the phylogenetic positions of *Darwinius* and *Afradapis* do not change (that is, adapiforms are placed as stem strepsirrhines, and *Darwinius* and *Afradapis* are placed within a clade of caenopithecines that forms the sister group of *Europolemur* and *Godinotia*). The ‘Shuanghuang petrosal’ described in ref. 47 is not scored as part of the *Eosimias sinensis* OTU in any of our analyses, following concerns of other authors⁴⁸ that this specimen might belong to one of several other small-bodied primate taxa known from the deposits that produced the fossil.

- Dagosto, M., Gebo, D. L. & Beard, K. C. Revision of the Wind River faunas, early Eocene of central Wyoming. Part 14. Postcranium of *Shoshonius cooperi* (Mammalia, Primates). *Ann. Carnegie Mus.* **68**, 175–211 (1999).
- Gebo, D. L., Dagosto, M., Beard, K. C. & Qi, T. Middle Eocene primate tarsals from China: implications for haplorhine evolution. *Am. J. Phys. Anthropol.* **116**, 83–107 (2001).
- Kay, R. F., Williams, B. A., Ross, C. F., Takai, M. & Shigehara, N. in *Anthropoid Origins: New Visions* (eds Ross, C. F. & Kay, R. F.) 91–135 (Kluwer, 2004).
- Masters, J. C. & Brothers, D. J. Lack of congruence between morphological and molecular data in reconstructing the phylogeny of the Galagonidae. *Am. J. Phys. Anthropol.* **117**, 79–93 (2002).
- Yoder, A. D. Relative position of the Cheirogaleidae in strepsirhine phylogeny: A comparison of morphological and molecular results and methods. *Am. J. Phys. Anthropol.* **94**, 25–46 (1994).
- Yoder, A. D., Irwin, J. A. & Payseur, B. A. Failure of the ILD to determine data combinability for slow loris phylogeny. *Syst. Biol.* **50**, 408–424 (2001).
- Franzen, J. L. Der sechste Messel-primat (Mammalia, Primates, Notharctidae, Cercamoniinae). *Senckenbergiana Lethaea* **80**, 289–303 (2000).
- Savage, R. J. G., Domning, D. P. & Thewissen, J. G. M. Fossil Sirenia of the west Atlantic and Caribbean region. V. The most primitive known sirenian, *Prorastomus sirenoides* Owen, 1855. *J. Vert. Paleol.* **14**, 427–449 (1994).
- Thewissen, J. G. M. & Williams, E. M. The early radiations of Cetacea (Mammalia): Evolutionary pattern and developmental correlations. *Annu. Rev. Ecol. Syst.* **33**, 73–90 (2002).
- Uhen, M. D. A new *Xenophorus*-like odontocete cetacean from the Oligocene of North Carolina and a discussion of the basal odontocete radiation. *J. Syst. Palaeontology* **6**, 433–452 (2008).
- Springer, M. S., Teeling, E. C., Madsen, O., Stanhope, M. J. & de Jong, W. W. Integrated fossil and molecular data reconstruct bat echolocation. *Proc. Natl Acad. Sci. USA* **98**, 6241–6246 (2001).
- Ciochon, R. L., Gingerich, P. D., Gunnell, G. F. & Simons, E. L. Primate postcrania from the late middle Eocene of Myanmar. *Proc. Natl Acad. Sci. USA* **98**, 7672–7677 (2001).
- Beard, K. C. *et al.* New sivaladapid primates from the Eocene Pondaung Formation of Myanmar and the anthropoid status of Amphipithecidae. *Bull. Carnegie Mus. Nat. Hist.* **39**, 67–76 (2007).
- Gunnell, G. F., Ciochon, R. L., Gingerich, P. D. & Holroyd, P. A. New assessment of *Pondaungia* and *Amphipithecus* (Primates) from the late middle Eocene of Myanmar, with a comment on ‘Amphipithecidae’. *Contr. Mus. Paleont. Univ. Mich.* **30**, 337–372 (2002).
- Beard, K. C. *et al.* Taxonomic status of purported primate frontal bones from the Eocene Pondaung Formation of Myanmar. *J. Hum. Evol.* **49**, 468–481 (2005).
- Ciochon, R. L. & Gunnell, G. F. Eocene primates from Myanmar: Historical perspectives on the origin of Anthropeidea. *Evol. Anthropol.* **11**, 156–168 (2002).
- MacPhee, R. D. E., Beard, K. C. & Qi, T. Significance of primate petrosal from middle Eocene fissure-fillings at Shuanghuang, Jiangsu Province, People’s Republic of China. *J. Hum. Evol.* **29**, 501–514 (1995).
- Ross, C. F. & Covert, H. H. The petrosal of *Omomyx carteri* and the evolution of the primate basicranium. *J. Hum. Evol.* **39**, 225–251 (2000).

LETTERS

Sleep deprivation impairs cAMP signalling in the hippocampus

Christopher G. Vecsey^{1,2}, George S. Baillie³, Devan Jaganath², Robbert Havekes², Andrew Daniels², Mathieu Wimmer^{1,2}, Ted Huang^{1,2}, Kim M. Brown³, Xiang-Yao Li⁴, Giannina Descalzi⁴, Susan S. Kim⁴, Tao Chen⁴, Yu-Ze Shang⁴, Min Zhuo⁴, Miles D. Houslay³ & Ted Abel²

Millions of people regularly obtain insufficient sleep¹. Given the effect of sleep deprivation on our lives, understanding the cellular and molecular pathways affected by sleep deprivation is clearly of social and clinical importance. One of the major effects of sleep deprivation on the brain is to produce memory deficits in learning models that are dependent on the hippocampus^{2–5}. Here we have identified a molecular mechanism by which brief sleep deprivation alters hippocampal function. Sleep deprivation selectively impaired 3', 5'-cyclic AMP (cAMP)- and protein kinase A (PKA)-dependent forms of synaptic plasticity⁶ in the mouse hippocampus, reduced cAMP signalling, and increased activity and protein levels of phosphodiesterase 4 (PDE4), an enzyme that degrades cAMP. Treatment of mice with phosphodiesterase inhibitors rescued the sleep-deprivation-induced deficits in cAMP signalling, synaptic plasticity and hippocampus-dependent memory. These findings demonstrate that brief sleep deprivation disrupts hippocampal function by interfering with cAMP signalling through increased PDE4 activity. Thus, drugs that enhance cAMP signalling

may provide a new therapeutic approach to counteract the cognitive effects of sleep deprivation.

To determine the cellular and molecular mechanisms by which a short 5-h period of sleep deprivation affects hippocampal function², we used *in vitro* electrophysiological recordings to examine the effects of sleep deprivation on hippocampal long-term potentiation (LTP), a form of synaptic plasticity that relies on molecular mechanisms that are also important for memory consolidation⁷. We examined several forms of NMDA (*N*-methyl-D-aspartate)-receptor-dependent LTP with different underlying molecular mechanisms to identify molecular targets of sleep deprivation. The long-term maintenance of LTP induced by 4 high-frequency trains of electrical stimuli applied at 5-minute intervals (spaced 4-train stimulation) or by 15 bursts of high-frequency stimuli applied at 5 Hz (theta-burst stimulation, TBS) depends on cAMP, PKA, transcription and translation^{8–10}. Both of these forms of LTP were impaired in hippocampal slices from mice that had been deprived of sleep for 5 h (Fig. 1a, b). The increased need for sleep that follows brief sleep deprivation dissipates in

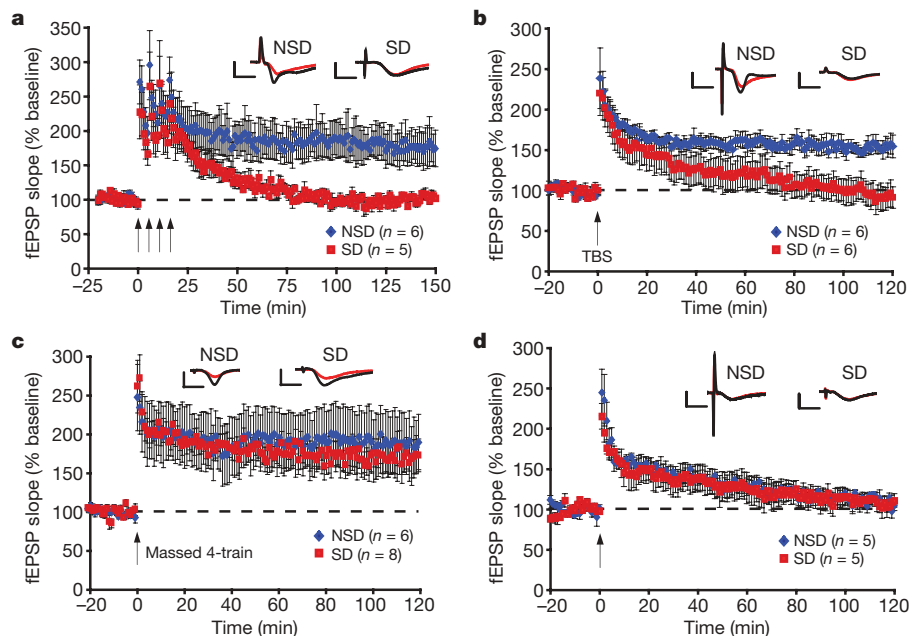


Figure 1 | Brief sleep deprivation specifically impairs forms of LTP that depend on the cAMP/PKA pathway. **a**, The maintenance of spaced 4-train LTP was significantly disrupted in slices from sleep-deprived (SD) mice ($P = 0.03$). fEPSPs, field excitatory postsynaptic potentials; NSD, non-sleep-

deprived. **b**, A similar deficit was observed in LTP induced by TBS ($P = 0.003$). **c**, Massed 4-train LTP was unaffected in hippocampal slices from sleep-deprived mice ($P = 0.67$). **d**, One-train LTP was unaffected in hippocampal slices from sleep-deprived mice ($P = 0.97$). All error bars denote s.e.m.

¹Neuroscience Graduate Group, ²Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³Neuroscience and Molecular Pharmacology, Wolfson and Davidson Buildings, Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK. ⁴Department of Physiology, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada.

approximately 2.5 h¹¹, and we found that the deficit in spaced 4-train LTP also recovered with 2.5 h of rest after 5 h of sleep deprivation (Supplementary Fig. 4). LTP induction by 4 high-frequency trains of stimuli applied at 5-second intervals (massed 4-train LTP), which produces a stable form of LTP that depends on translation¹² but does not require cAMP/PKA signalling^{12,13}, was unimpaired by sleep deprivation (Fig. 1c). One-train LTP, a cAMP/PKA-independent form of LTP⁶ was also unaffected by sleep deprivation (Fig. 1d). The lack of an effect of sleep deprivation on massed 4-train LTP and 1-train LTP suggests that brief sleep deprivation does not affect molecular mechanisms that are required for the induction and expression of these forms of LTP, such as NMDA-receptor activation, Ca²⁺ influx and activation of Ca²⁺-calmodulin dependent kinase II (CaMKII)^{8,14,15}. Because massed 4-train requires translation, the fact that sleep deprivation does not affect this form of LTP suggests that brief sleep deprivation does not generally disrupt translational processes, but instead specifically alters mechanisms that depend upon cAMP/PKA signalling. Whole-cell recordings from area CA1 confirmed that NMDA-receptor function was unaffected by sleep deprivation (Supplementary Fig. 3). These results contrast with studies using longer periods of sleep deprivation, or sleep deprivation that involves exploration of a new environment—both of which affect the initial induction of LTP as well as NMDA-receptor function^{4,16–20}. We also did not observe any effects of sleep deprivation on basal synaptic properties or short-term plasticity (Supplementary Fig. 2), suggesting that the disruption of spaced 4-train and theta-burst LTP is in fact due to disruption of signalling mechanisms underlying these forms of LTP, and not due to non-specific effects on hippocampal function.

Because of the role of cAMP signalling in 4-train and theta-burst LTP^{8,13}, we next assessed the effects of sleep deprivation on LTP induced by specific activation of the cAMP pathway using the adenylylase activator forskolin (FSK). The long-term maintenance of LTP induced by FSK treatment (50 μ M) was impaired in hippocampal slices from sleep-deprived mice (Fig. 2a). Using biochemical assays, we found that baseline cAMP levels were significantly reduced

in the CA1 region of hippocampal slices from sleep-deprived mice, as were cAMP levels induced by FSK treatment (Fig. 2c). These findings demonstrate that sleep deprivation limits the ability of cells within the CA1 region of the hippocampus to respond to adenylylase activation. Using immunohistochemistry, we found that phosphorylation at Ser 133 of CREB, one of the downstream targets of cAMP signalling in the hippocampus²¹, was decreased by sleep deprivation in the CA1 and dentate gyrus regions of the hippocampus (Supplementary Fig. 7). Consistent with our published findings that brief sleep deprivation selectively impairs hippocampal function², no effect of sleep deprivation on CREB phosphorylation was observed in the amygdala (Supplementary Fig. 7). This demonstrates that deficient cAMP signalling due to sleep deprivation impairs the activation of at least one major downstream target (CREB) *in vivo*. Cyclic nucleotide phosphodiesterases (PDEs) are enzymes that provide the sole route for cAMP degradation in cells²². We found that co-treatment with 3-isobutyl-1-methylxanthine (IBMX; 30 μ M), a broad-spectrum inhibitor of PDEs, rescued FSK-induced LTP as well as cAMP induction in hippocampal slices from sleep-deprived mice (Fig. 2b, c). These findings demonstrate that reversing deficits in cAMP signalling ameliorates the deficits in FSK-induced LTP, and suggest that PDE activity might be increased by sleep deprivation.

The cAMP-specific PDE4 has a major role in regulating cAMP signalling in the brain²². Therefore, we tested whether sleep deprivation increased PDE4 levels and/or activity in the hippocampus. We found that PDE4-specific cAMP breakdown was increased in sleep-deprived mice (Fig. 3a), whereas non-PDE4 activity was unaffected

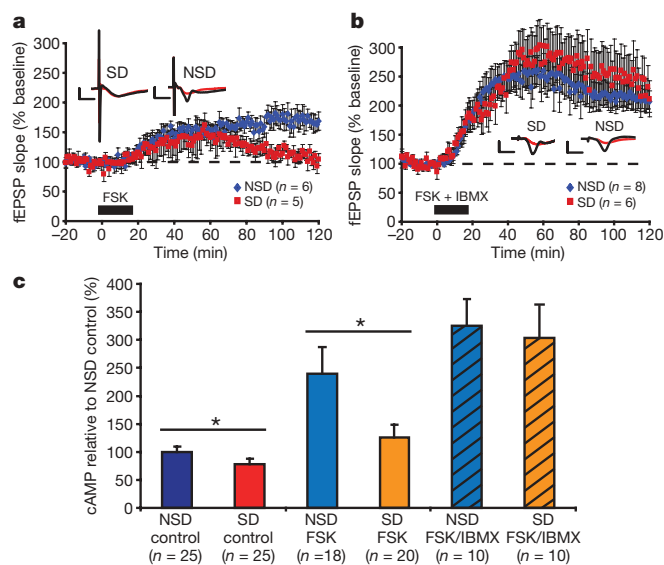


Figure 2 | PDE inhibition rescues impairments in FSK-induced cAMP levels and LTP produced by brief sleep deprivation. **a**, LTP induced by the adenylylase activator FSK was impaired in sleep-deprived mice relative to controls ($P = 0.007$). **b**, LTP induced by co-treatment with FSK and the PDE inhibitor IBMX was unaffected by sleep deprivation ($P = 0.48$). **c**, Sleep deprivation decreased baseline cAMP levels in CA1 regions of vehicle-treated slices ($P = 0.02$) and significantly reduced FSK-induced cAMP levels ($P = 0.04$). Co-application of FSK and IBMX resulted in similar cAMP levels in CA1 regions from sleep-deprived and non-sleep-deprived mice ($P = 0.82$). All error bars denote s.e.m.

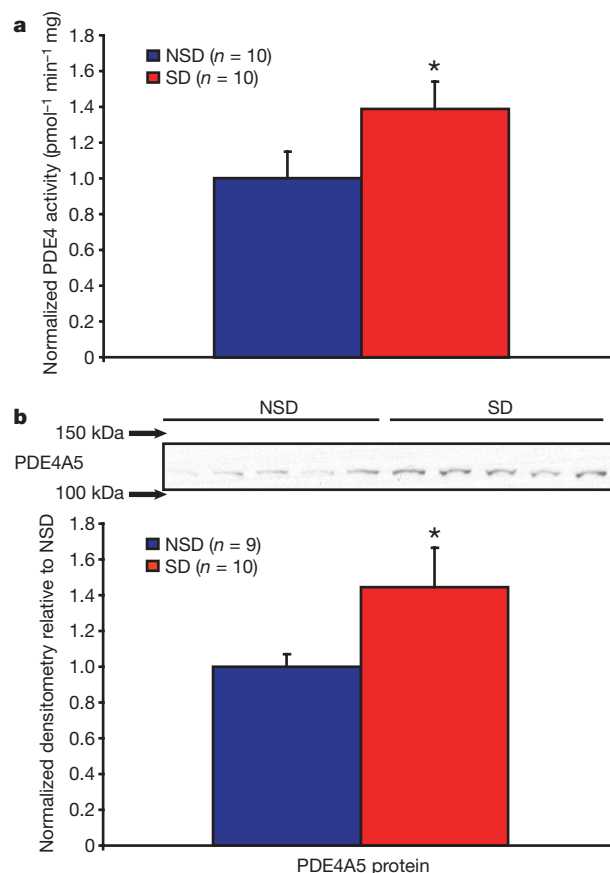


Figure 3 | Sleep deprivation increases PDE4 activity and protein levels in the hippocampus. **a**, PDE4 activity was significantly upregulated in hippocampi from sleep-deprived compared with non-sleep-deprived mice ($P = 0.039$). **b**, The PDE4 isoform PDE4A5 was significantly upregulated by sleep deprivation in the hippocampus ($P = 0.033$). A sample blot is shown, with the nearest size markers indicated by arrows. All error bars denote s.e.m.

(sleep-deprived: 50.6 ± 2.1 pmol $\text{mg}^{-1} \text{min}^{-1}$, non-sleep-deprived: 53.4 ± 1.9 pmol $\text{mg}^{-1} \text{min}^{-1}$, $P = 0.33$). In mice and humans, PDE4 isoforms are generated by four genes: *Pde4a-d*²². Immunoblotting with a pan-PDE4A antibody showed a trend in sleep-deprived hippocampal tissue towards an increase in a 125-kilodalton (kDa) species, which is of a size expected for the long PDE4A5, PDE4A10 and PDE4A11 isoforms²³ (Supplementary Fig. 6). Probing our hippocampal samples with antisera specific for each of these isoforms showed that PDE4A5 protein was significantly increased by sleep deprivation in the hippocampus, whereas PDE4A10 and PDE4A11 were not detected (Fig. 3b for PDE4A5; data not shown for null results with PDE4A10 and PDE4A11). Interestingly, transcripts for PDE4A5 are selectively expressed in the brain with highest levels seen in hippocampal areas CA1 and dentate gyrus²⁴, the two regions in which reduction in phosphorylated-CREB were observed after sleep deprivation (Supplementary Fig. 7). We also found that sleep deprivation elicited upregulation of *Pde4* gene expression, although the pattern of changes was different between transcript and protein levels, suggesting that translation regulation mechanisms may be involved (Supplementary Fig. 5).

To test whether increased PDE4 activity is responsible for deficits in hippocampal function induced by sleep deprivation, we first examined whether the PDE4-selective inhibitor rolipram²² could rescue deficits in LTP induced by sleep deprivation. Indeed, rolipram treatment (0.1 μM) during spaced 4-train stimulation rescued the maintenance of this form of LTP in slices from sleep-deprived mice. This dose of rolipram had no effect on 4-train LTP in slices from control mice, suggesting that this result was not simply due to a non-specific enhancement of LTP (Fig. 3a, b). To rule out the possibility of a ceiling effect, we assessed the effects of rolipram in slices from sleep-deprived and control mice on 1-train LTP, which is normally a transient form of plasticity. Here, rolipram enhanced the maintenance of 1-train LTP in control mice. However, rolipram had no significant effect in sleep-deprived mice, demonstrating that sleep-deprived mice have a reduced sensitivity to the effects of PDE4 inhibition (Supplementary Fig. 8). These findings suggest that sleep deprivation makes it more difficult to reach the threshold for induction of cAMP/PKA-dependent LTP, consistent with our observation that sleep deprivation increases the activity of a specific PDE4 isoform that is known to set the threshold for activation when recruited to specific signalling complexes²⁵.

We next assessed whether *in vivo* PDE4 inhibition could prevent impairments in hippocampus-dependent memory consolidation induced by post-training sleep deprivation. Immediately after single-trial contextual fear conditioning, mice were either deprived of sleep by gentle handling for 5 h or left undisturbed in their home cages. Mice received two intraperitoneal injections of either rolipram (1 mg kg^{-1}) or vehicle, immediately and 2.5 h after training. The timings of these injections were based on data showing that the effects of rolipram on hippocampal cAMP levels last for about 3 h (personal communication, J. O'Donnell). Memory was assessed using a 5-min retrieval test in the trained context 1 day after conditioning and in an altered context 2 days after conditioning. Context-specific memory, a measure that is particularly dependent on hippocampal function²⁶, was measured by subtracting the percentage freezing in the altered context from the percentage freezing in the trained context (see Supplementary Fig. 9 for data from these two tests). Sleep deprivation significantly impaired context-specific memory, and rolipram treatment rescued this deficit (Fig. 4c). The ability of rolipram to rescue deficits in hippocampus-dependent memory produced by sleep deprivation demonstrates that disruption of cAMP signalling is responsible for the effects of sleep deprivation *in vivo*.

A major challenge in the field of sleep research has been to determine how the sleep disruptions associated with neurological and psychiatric disorders, ageing and everyday living affect cognitive function. The findings presented here define a molecular mechanism underlying the effects of sleep deprivation on hippocampal function

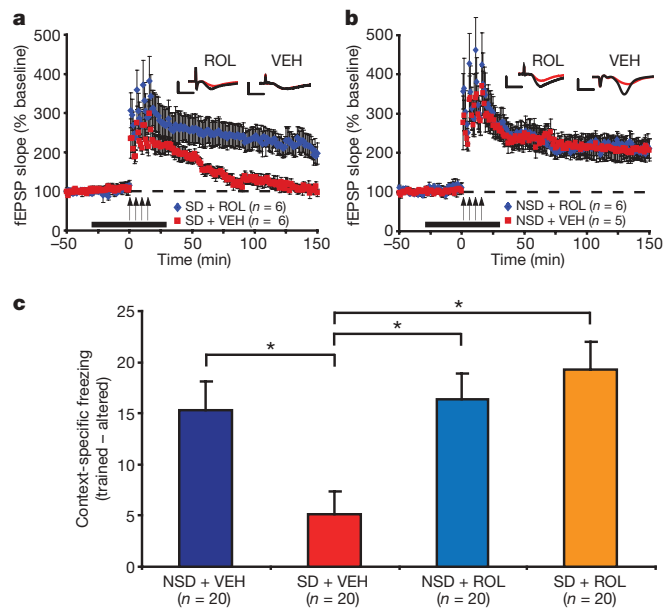


Figure 4 | The PDE4 inhibitor rolipram rescues LTP and memory deficits caused by sleep deprivation. **a**, Rolipram (ROL) treatment rescued deficits in spaced 4-train LTP due to sleep deprivation ($P = 0.003$). VEH, vehicle. **b**, However, rolipram showed no further enhancement of spaced 4-train LTP in non-sleep-deprived mice ($P = 1.0$). The black bar in **a** and **b** represents the time of rolipram treatment. **c**, Sleep deprivation significantly impaired context-specific memory ($P = 0.02$), and treatment with rolipram rescued this deficit ($P = 0.0009$) without affecting memory in non-sleep-deprived mice ($P = 0.99$). All error bars denote s.e.m.

at the behavioural and synaptic level (Supplementary Fig. 1), and they lay the groundwork for further analysis of the functional biochemistry of sleep deprivation and the development of new therapeutics to ameliorate the effect of sleep deprivation. These experiments using the PDE4 inhibitor rolipram are the first, to our knowledge, to rescue synaptic plasticity and memory deficits produced by a brief period of sleep deprivation. Our data suggest that if compounds can be developed that block the activity, expression or targeting of the PDE4A5 isoform, they may prove useful in the treatment of the cognitive effects of sleep deprivation. Notably, circadian oscillation of cAMP in the hippocampus has recently been linked to the persistence of memory²⁷, so such drugs may also be useful in treating memory deficits associated with alterations in circadian rhythms. It has proven difficult to generate active-site-directed inhibitors showing selectivity for each of the four PDE4 sub-families because of the highly conserved structure of their catalytic unit²⁵, but it may be possible to develop compounds that disrupt the targeting of specific PDE4 isoforms from appropriate scaffolds in cells^{22,25}.

METHODS SUMMARY

C57BL/6J male mice (2–5 months of age) were housed individually on a 12 h/12 h light/dark schedule with lights on at 7:00 (ZT0) and handled for 6 days. Mice were sleep-deprived in their home cages for 5 h by gentle handling beginning at ZT5 or left undisturbed (non-sleep-deprived mice). For contextual fear conditioning experiments, animals were placed in a new chamber for 3 min, and received a 2-s, 1.5 mA footshock after 2.5 min. Half of the mice were deprived of sleep for 5 h post-training. Mice received intraperitoneal injections of rolipram (1 mg kg^{-1}) or vehicle (2% dimethylsulphoxide (DMSO) in 0.9% saline) immediately and 2.5 h after training. Testing of contextual memory was performed 24 h after training in the trained context and 48 h after training in a new chamber.

Electrophysiological recordings were carried out as previously reported²⁸. One-train LTP was induced by a single 100-Hz, 1-s duration train of stimuli. Four-train LTP consisted of four trains applied with a 5-min intertrain interval; for massed 4-train LTP a 5-s intertrain interval was used. TBS consisted of 40-ms duration, 100 Hz bursts delivered at 5 Hz for 3 s (15 bursts of 4 pulses per burst, for a total of 60 pulses). Chemical LTP was induced by treatment of slices for 15 min with 5 μM FSK in 0.1% ethanol, or a combination of 50 μM FSK and

30 μ M IBMX (in water). Rolipram (0.1 μ M in 0.1% DMSO) was applied for 60 min, beginning 30 min before tetanization.

cAMP assays on CA1 regions of hippocampal slices 10 min after treatment for 15 min with FSK (50 μ M), FSK plus IBMX (30 μ M), or vehicle (0.1% ethanol) were performed by radioimmunoassay according to kit instructions. cAMP-specific PDE activity assays²⁹ and western blots for PDE4A5 (ref. 30) were performed as previously described.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 May; accepted 1 September 2009.

- Hublin, C., Kaprio, J., Partinen, M. & Koskenvuo, M. Insufficient sleep—a population-based study in adults. *Sleep* **24**, 392–400 (2001).
- Graves, L. A., Heller, E. A., Pack, A. I. & Abel, T. Sleep deprivation selectively impairs memory consolidation for contextual fear conditioning. *Learn. Mem.* **10**, 168–176 (2003).
- Guan, Z., Peng, X. & Fang, J. Sleep deprivation impairs spatial memory and decreases extracellular signal-regulated kinase phosphorylation in the hippocampus. *Brain Res.* **1018**, 38–47 (2004).
- McDermott, C. M. et al. Sleep deprivation causes behavioral, synaptic, and membrane excitability alterations in hippocampal neurons. *J. Neurosci.* **23**, 9687–9695 (2003).
- Smith, C. & Rose, G. M. Evidence for a paradoxical sleep window for place learning in the Morris water maze. *Physiol. Behav.* **59**, 93–97 (1996).
- Nguyen, P. V. & Woo, N. H. Regulation of hippocampal synaptic plasticity by cyclic AMP-dependent protein kinases. *Prog. Neurobiol.* **71**, 401–437 (2003).
- Martin, S. J., Grimwood, P. D. & Morris, R. G. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
- Huang, Y. Y. & Kandel, E. R. Recruitment of long-lasting and protein kinase A-dependent long-term potentiation in the CA1 region of hippocampus requires repeated tetanization. *Learn. Mem.* **1**, 74–82 (1994).
- Nguyen, P. V., Abel, T. & Kandel, E. R. Requirement of a critical period of transcription for induction of a late phase of LTP. *Science* **265**, 1104–1107 (1994).
- Nguyen, P. V. & Kandel, E. R. Brief theta-burst stimulation induces a transcription-dependent late phase of LTP requiring cAMP in area CA1 of the mouse hippocampus. *Learn. Mem.* **4**, 230–243 (1997).
- Huber, R., Deboer, T. & Tobler, I. Effects of sleep deprivation on sleep and sleep EEG in three mouse strains: empirical data and simulations. *Brain Res.* **857**, 8–19 (2000).
- Scharf, M. T. et al. Protein synthesis is required for the enhancement of long-term potentiation and long-term memory by spaced training. *J. Neurophysiol.* **87**, 2770–2777 (2002).
- Woo, N. H., Duffy, S. N., Abel, T. & Nguyen, P. V. Temporal spacing of synaptic stimulation critically modulates the dependence of LTP on cyclic AMP-dependent protein kinase. *Hippocampus* **13**, 293–300 (2003).
- Morris, R. G., Anderson, E., Lynch, G. S. & Baudry, M. Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature* **319**, 774–776 (1986).
- Lynch, G., Kessler, M., Halpain, S. & Baudry, M. Biochemical effects of high-frequency synaptic activity studied with *in vitro* slices. *Fed. Proc.* **42**, 2886–2890 (1983).
- Campbell, I. G., Guinan, M. J. & Horowitz, J. M. Sleep deprivation impairs long-term potentiation in rat hippocampal slices. *J. Neurophysiol.* **88**, 1073–1076 (2002).
- Chen, C., Hardy, M., Zhang, J., LaHoste, G. J. & Bazan, N. G. Altered NMDA receptor trafficking contributes to sleep deprivation-induced hippocampal synaptic and cognitive impairments. *Biochem. Biophys. Res. Commun.* **340**, 435–440 (2006).
- Kopp, C., Longordo, F., Nicholson, J. R. & Luthi, A. Insufficient sleep reversibly alters bidirectional synaptic plasticity and NMDA receptor function. *J. Neurosci.* **26**, 12456–12465 (2006).
- McDermott, C. M., Hardy, M. N., Bazan, N. G. & Magee, J. C. Sleep deprivation-induced alterations in excitatory synaptic transmission in the CA1 region of the rat hippocampus. *J. Physiol. (Lond.)* **570**, 553–565 (2006).
- Tartar, J. L. et al. Hippocampal synaptic plasticity and spatial learning are impaired in a rat model of sleep fragmentation. *Eur. J. Neurosci.* **23**, 2739–2748 (2006).
- Josselyn, S. A. & Nguyen, P. V. CREB, synapses and memory disorders: past progress and future challenges. *Curr. Drug Targets CNS Neurol. Disord.* **4**, 481–497 (2005).
- Houslay, M. D. & Adams, D. R. PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. *Biochem. J.* **370**, 1–18 (2003).
- Wallace, D. A. et al. Identification and characterization of PDE4A11, a novel, widely expressed long isoform encoded by the human PDE4A cAMP phosphodiesterase gene. *Mol. Pharmacol.* **67**, 1920–1934 (2005).
- McPhee, I., Cochran, S. & Houslay, M. D. The novel long PDE4A10 cyclic AMP phosphodiesterase shows a pattern of expression within brain that is distinct from the long PDE4A5 and short PDE4A1 isoforms. *Cell. Signal.* **13**, 911–918 (2001).
- Houslay, M. D., Schafer, P. & Zhang, K. Y. Keynote review: phosphodiesterase-4 as a therapeutic target. *Drug Discov. Today* **10**, 1503–1519 (2005).
- Frankland, P. W., Cestari, V., Filipkowski, R. K., McDonald, R. J. & Silva, A. J. The dorsal hippocampus is essential for context discrimination but not for contextual conditioning. *Behav. Neurosci.* **112**, 863–874 (1998).
- Eckel-Mahan, K. L. et al. Circadian oscillation of hippocampal MAPK activity and cAMP: implications for memory persistence. *Nature Neurosci.* **11**, 1074–1082 (2008).
- Vecsey, C. G. et al. Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB:CBP-dependent transcriptional activation. *J. Neurosci.* **27**, 6128–6140 (2007).
- Lobban, M., Shakur, Y., Beattie, J. & Houslay, M. D. Identification of two splice variant forms of type-IVB cyclic AMP phosphodiesterase, DPD (rPDE-IVB1) and PDE-4 (rPDE-IVB2) in brain: selective localization in membrane and cytosolic compartments and differential expression in various brain regions. *Biochem. J.* **304**, 399–406 (1994).
- Huston, E. et al. The cAMP-specific phosphodiesterase PDE4A5 is cleaved downstream of its SH3 interaction domain by caspase-3. Consequences for altered intracellular distribution. *J. Biol. Chem.* **275**, 28063–28074 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Sehgal and J. Hawk for their comments on the manuscript. We thank J. O'Donnell for his help with planning rolipram treatment experiments. We thank J. Bibb for discussions and comments on the manuscript. We thank S. Fluharty and J. Lindstrom for the use of their gamma scintillation counters, and T. Bale, N. Goel, K. Semsar and S. Teegarden for their help with corticosterone assays. We also thank P. Hernandez, N. Khatib, A. Park, J. Lederman, C. Florian and W. Lu for their help with experiments. This research was supported by Systems and Integrative Biology Training grant GM07517 (to C.G.V., M. Nusbaum PI), NIH training grant HL07953 (to C.G.V., A. I. Pack PI), the Netherlands Organization for Scientific Research NWO-Rubicon grant 825.07.029 (to R.H.), the National Institutes of Health (AG017628; to T.A., A. I. Pack PI), SCOR grant HL060287 (to T.A., A. I. Pack PI), HFSP grant RGSP/2005 (to T.A.), the Medical Research Council (UK) grant G0600765 (to M.D.H. and G.S.B.), the European Union grant LSHB-CT-2006-037189 (to M.D.H.), the Fondation Leducq grant 06CVD02 (to M.D.H. and G.S.B.), CIHR84256 (to M.Z.), and a UK Engineering and Physical Sciences Research Council training grant (to K.M.B.).

Author Contributions Experiments were conceived and designed by C.G.V., T.A., G.S.B., M.D.H. and M.Z. Behavioural and gene expression experiments were carried out by D.J. and C.G.V. Electrophysiological recordings were carried out by C.G.V. and T.H. cAMP assays were carried out by C.G.V. Western blots were carried out by K.M.B. PDE activity assays were carried out by G.S.B. Immunohistochemistry experiments were carried out by R.H. and A.D. EEG/EMG recordings were carried out by M.W. Sleep deprivation and electrophysiology for whole-cell patch-clamp recordings were performed by X.-Y.L., G.D., S.S.K., T.C. and Y.-Z.S. Manuscript was prepared by C.G.V. and T.A., with input from D.J., R.H., M.W., G.S.B. and M.D.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.A. (abele@sas.upenn.edu).

METHODS

Subjects. C57BL/6J adult male mice (2–4 months of age) were housed individually on a 12 h/12 h light/dark schedule with lights on at 7:00 (ZT0). For electrophysiological analysis of NMDA-receptor function, mice were studied at 6 weeks of age. Food and water were available *ad libitum* throughout the experiment. Each animal was handled daily for 6 days before sleep deprivation. For electrophysiological, biochemical and gene expression experiments, starting at approximately ZT5, mice were sleep-deprived (SD) in their home cages for 5 h by gentle handling³¹. Non-sleep-deprived (NSD) mice were left undisturbed in their home cages. Using electroencephalographic (EEG) and electromyographic (EMG) recordings, we found that this form of sleep deprivation resulted in ~95% wakefulness in SD mice, compared with ~35% wakefulness during the same time period in NSD mice (data not shown). All experiments were approved by the Institution of Animal Care and Use Committee of the University of Pennsylvania, or by the Animal Studies Committee at the University of Toronto, and were carried out in accordance with all National Institutes of Health guidelines.

Behaviour. Foreground fear conditioning experiments were performed as previously described³². Mice were acclimated to the training/testing room during 6 days of handling. Starting at ZT4–5, mice were placed into the conditioning chamber (Med Associates) and received a 2-s, 1.5-mA scrambled footshock 2.5 min after placement into the chamber. Mice were removed from the chamber 30 s after the shock. For contextual fear conditioning experiments, animals were placed in a new chamber (Med Associates) for 3 min, and received a 2-s, 1.5-mA footshock after 2.5 min. Half of the mice were deprived of sleep as described earlier for 5 h after training. Mice received two intraperitoneal injections of rolipram (1 mg kg⁻¹) or vehicle (2% DMSO in 0.9% saline) immediately and 2.5 h after training. For trained context testing, mice received one 5-min exposure to the same conditioned context in the absence of shock 24 h after conditioning. For altered context testing, mice received one 5-min exposure to a new context (another conditioning chamber with smooth flat floor, altered dimensions, and a new odorant), 48 h after conditioning. Conditioning was assayed by measuring freezing behaviour, the complete absence of movement³³. Freezing was scored during conditioning as well as testing. The behaviour of each mouse was sampled at 5-s intervals, and the percentage of those intervals in which the mouse froze was calculated. Context-specific freezing was calculated by subtracting the freezing level in the altered context from the freezing level in the trained context for each animal.

Electrophysiology. Immediately after sleep deprivation, SD and time-matched NSD mice were killed by cervical dislocation, and their hippocampi were dissected rapidly in iced oxygenated artificial cerebrospinal fluid (ACSF). Field-potential recordings were carried out as previously reported³². In brief, 400 μ m thick transverse hippocampal slices were placed in an interface chamber³⁴ and continuously perfused with oxygenated ACSF while they equilibrated for at least 1.5 h at 30 °C before starting electrophysiological recordings. A bipolar stimulating electrode (A-M systems, Inc.; 0.002 inches diameter nichrome wire) placed in the stratum radiatum was used to elicit action potentials in the Schaffer collateral pathway. An ACSF-filled glass microelectrode (A-M systems, Inc.; 1.5 mm \times 0.85 mm) with a resistance between 0.5 and 3 M Ω placed in the stratum radiatum region of CA1 was used to record the resulting fEPSPs. Data were acquired using Clampex 8.2 (Axon Instruments, Inc.), and analysed using Clampfit 8.2 (Axon Instruments, Inc.). Peak fEPSP amplitude was required to be at least 3 mV, and stimulus intensity was set to produce 40% of the maximal response. Stimulations occurred every minute, and a 20-min baseline period was recorded in each experiment, and recordings continued for at least 2 h after LTP induction. Initial fEPSP slopes were normalized against the average of the 20 baseline traces.

One-train LTP was induced by a single 100 Hz, 1-s duration train of stimuli. This form of LTP has been shown to be dependent on the NMDA receptor³⁵, calcium influx and CaMKII activation, but is independent of PKA or translation^{36,37}. Four-train LTP, an NMDA-receptor-dependent form of LTP³⁸ consisted of four such trains applied with a 5-min intertrain interval. The long-term maintenance of this form of LTP has a further reliance on cAMP, PKA, transcription and translation^{36,39}. Massed 4-train LTP also used four such trains of stimuli, but with a 5-s intertrain interval. This long-lasting form of LTP, despite being dependent on the NMDA receptor and translation, does not require PKA³⁸. TBS consisted of 40-ms duration, 100-Hz bursts delivered at 5 Hz for 3 s (15 bursts of 4 pulses per burst, for a total of 60 pulses). The induction of this form of LTP depends on the NMDA receptor, and its maintenance depends on cAMP⁴⁰. Chemical LTP was induced by treatment of slices for 15 min with 50 μ M molecular grade FSK, an adenylate cyclase activator⁴¹ (Sigma-Aldrich) in 0.1% ethanol, or a combination of 50 μ M FSK and the broad-spectrum PDE inhibitor 30 μ M IBMX⁴² (dissolved in water; Tocris Cookson, Inc.). The cAMP-specific PDE4 inhibitor rolipram⁴³ (Sigma-Aldrich) was applied at 0.1 μ M in 0.1%

DMSO (Sigma-Aldrich) for 60 min, beginning 30 min before tetanization. In all LTP figures, representative sample sweeps are shown for each LTP experiment. The red sweeps represent the average of the first five baseline sweeps, and the black sweeps represent the last five sweeps in the recording. All scale bars are 5 mV/5 ms.

Input–output characteristics were examined by recording fEPSPs in CA1 resulting from stimuli of increasing intensity. Initial fEPSP slopes were plotted against the corresponding presynaptic fibre volley amplitudes, and the resulting plots were fit with linear regressions. The slope of the maximum elicited fEPSP in each slice was also recorded. Paired-pulse facilitation, a short-term form of synaptic plasticity, was measured in slices from sleep-deprived and control mice. Pairs of stimuli were delivered with varying delays (300, 200, 100, 50 and 25 ms) between the two stimuli, and the initial fEPSP slope from the second stimulus was plotted relative to the slope from the first stimulus.

For whole-cell patch-clamp recordings in slices from SD and NSD mice, coronal brain slices (350 μ m) containing the hippocampus were prepared as previously described⁴⁴. Slices were kept in a submerged recovery chamber at room temperature for at least 1 h before electrophysiological experiments, in oxygenated (95% O₂ and 5% CO₂) ACSF containing the following (in mM): 124 NaCl, 2.5 KCl, 2 CaCl₂, 2 MgSO₄, 25 NaHCO₃, 1 NaH₂PO₄ and 10 glucose. Whole-cell patch-clamp recording experiments were performed in a recording chamber on the stage of an Olympus BX51WI microscope with infrared differential interference contrast (DIC) optics for visualization of whole-cell patch-clamp recordings. The recording pipettes (3–5 M Ω) were filled with solution containing (in mM): 102 CsMeSO₃, 3.7 NaCl, 5 QX-314 chloride, 5 TEA-Cl, 10 BAPTA, 0.2 EGTA, 20 HEPES, 0.3 Mg-ATP and 0.3 Na₃-GTP (adjusted to pH 7.2 with CsOH). Excitatory postsynaptic currents (EPSCs) were recorded from hippocampal CA1 pyramidal neurons with an Axon 200B amplifier (Molecular Devices), and the stimulations were delivered by a bipolar tungsten stimulating electrode placed in striatum radiatum. Data were collected and analysed using pClamp 9.2 software. Picrotoxin (100 μ M) was always present to block GABA_A (γ -aminobutyric acid subtype A) receptor-mediated inhibitory synaptic currents. AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor-mediated EPSCs were induced by repetitive stimulations at 0.05 Hz, and neurons were voltage-clamped at -70 mV (liquid junction potential corrected). The NMDA-receptor-mediated EPSCs were recorded in the presence of the AMPA-receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) (20 μ M) and evoked at 0.05 Hz. Neurons were voltage-clamped at 0 mV for the experiments examining the input–output relationship of NMDA-receptor-mediated EPSCs. Data were discarded if access resistance changed by more than 15% during experiments.

Biochemistry. For cAMP assays, hippocampal slices were prepared as for electrophysiology, and were treated for 15 min with FSK (50 μ M), FSK plus IBMX (30 μ M), or vehicle. At 10 min after drug treatment, slices were removed from the rig and the CA1 region was dissected and flash-frozen in a dry-ice and ethanol slurry. Samples were stored at -80 °C until enough had been accumulated to run the assay. For the assay itself, a cAMP-[¹²⁵I]-radioimmunoassay kit was used (PerkinElmer). Tissue was homogenized on ice in 6% TCA, a set amount (~30,000 c.p.m.) of tritiated cAMP (adenosine 3', 5'-cyclic phosphate, ammonium salt, [2,8-³H]; PerkinElmer) was added to each sample, and extracts were centrifuged at 2,500g at 4 °C for 15 min. The pellet was frozen and used later in Bradford assays to determine the total protein concentration in each sample. After extracting the supernatants four times with water-saturated ether, samples were evaporated in a speed-vac and then resuspended in assay buffer. Recovery was assessed by measuring the amount of [³H]-cAMP in each sample using a scintillation counter. The assay was run according to kit instructions for non-acetylated procedure, and the reaction was run at 4 °C overnight. [¹²⁵I] was measured using a gamma scintillation counter. On the basis of Bradford assays, recovery assessment and standard curves, cAMP levels (pmol mg⁻¹ protein) were calculated for each sample. Because of interassay variability, raw data was normalized for all groups relative to the NSD control average for each assay. Data are therefore presented as percentage NSD control cAMP level.

For PDE activity assays, hippocampi were dissected from SD and NSD mice and immediately frozen in a dry-ice and ethanol slurry. The cAMP-specific PDE activity assay was adapted from the radioassay procedure described previously⁴⁵, as detailed in ref. 46. In this two-step procedure the samples to be assayed are (in the first step) incubated with 1 μ M 8-[³H]-labelled cAMP substrate for 10 min at 30 °C, in the absence or presence of rolipram (10 μ M) to determine the rolipram-inhibited PDE4 activity. In the second step, the [³H]-labelled cAMP product of cAMP hydrolysis, 5' AMP, was dephosphorylated to adenosine by incubation with 0.2 mg ml⁻¹ snake venom. The negatively charged unhydrolysed cAMP was then separated from the uncharged adenosine by incubation with Dowex ion exchange resin (Dowex-1-chloride); this removes the charged nucleotides but not the uncharged nucleosides. The amount of unbound [³H]-adenosine in the

supernatant was determined by scintillation counting to calculate the rate of cAMP hydrolysis. PDE4-specific activity was calculated by subtracting the activity found in the presence of rolipram from the activity in its absence.

For PDE4 western blots, hippocampi were flash-frozen on a dry-ice and ethanol slurry and stored at -80°C . One-hundred-and-fifty-four micrograms of total protein was resolved using NuPAGE 4–12% Bis-Tris gel and NuPAGE MOPS SDS running buffer (Invitrogen) for 1 h at 190 V. The separated proteins were transferred to nitrocellulose membrane (Protran, Whatman GmbH) for 1 h at 25 V using NuPAGE transfer buffer. The membranes were blocked in 5% marvel/TBST (TBS containing 0.05% (v/v) Tween 20) for at least 1 h with gentle shaking. Antibodies specific for PDE4A, PDE4B, PDE4D, PDE4A5, PDE4A10 and PDE4A11 (1:1,000 for each)^{47,48} or α -tubulin (1:5,000, mouse) (Sigma-Aldrich) were added in 1% marvel/TBST and incubated at 4°C overnight. The membranes were washed 3 times for 10 min in TBST. Peroxidase-conjugated anti-rabbit (Sigma-Aldrich) or anti-mouse (GE Healthcare UK Ltd) were added 1:2,000 in 1% marvel/TBST and incubated for 1 h at room temperature. Membranes were washed as before, then incubated in Pierce ECL Western Blotting Substrate (Thermo Scientific) for 30 s and the signal was detected using film (Kodak, Carestream Health, Inc.). Densitometry was performed using Quantity One software (BioRad Laboratories).

For phosphorylated-CREB immunohistochemistry, mice were anaesthetized with isoflurane and perfused with PBS followed by 4% paraformaldehyde in PBS. Fixed brains were dissected and then cryoprotected in 30% sucrose in PBS overnight at 4°C . Brains were mounted on cryostat chucks using OCT. Coronal sections were cut at a thickness of 25 μm and stored in PBS with 0.1% sodium azide. After rinsing in PBS (3 \times), incubation in 0.3% H_2O_2 (30 min), and rinsing again in PBS (3 \times), sections were preincubated in 5% normal goat serum (Jackson ImmunoResearch Laboratories) in PBS for 20 min at room temperature. After preincubation, sections were incubated with rabbit-anti-pCREB IgG (1:1,500; Upstate Cell Signaling Solutions) in 1% normal goat serum, 0.3% Triton X-100 and 0.1% azide in PBS at room temperature for 24 h. Sections were then rinsed in PBS (3 \times) and incubated for 3 h with biotinylated goat-anti-rabbit IgG (1:500; Jackson ImmunoResearch Laboratories) in PBS containing 0.3% Triton X-100 and 1% normal goat serum at 5°C . After rinsing with PBS (3 \times), sections were incubated with the avidin-biotin-horseradish peroxidase complex (1:500 ABC kit, Vector Laboratories) in PBS containing 0.3% Triton X-100 for 2 h at room temperature. Finally, after rinsing in PBS for 6 h, sections were processed with diaminobenzidine (0.02% in Tris-HCl, pH 7.6) with 100 μl 0.1% H_2O_2 as a reaction initiator. After 7 min, the reaction was terminated by washing the sections with PBS. Sections were mounted with gelatin (0.7%) and dried for 24 h. The mounted sections were then dehydrated using ethanol and xylene, coverslipped with DPX (Merck), and dried for 24 h. Relative optical densities of pCREB immunoreactivity were quantified in subregions of the dorsal hippocampus (three sections per animal, averaged for both hemispheres) and in the amygdala (3–8 amygdalae) using ImageJ image analysis software. The optical density is expressed in arbitrary units corresponding to grey levels. The optical density values for each region were normalized to the background labelling within the same region to normalize for variability in staining between sections.

For corticosterone assays, animals were killed by cervical dislocation in control mice or mice that had been sleep-deprived for 5 h, and trunk blood was collected. After 30–60 min at room temperature, samples were centrifuged at 2,300g for 10 min to isolate the supernatant (plasma), which was then stored at -80°C until enough samples had been accumulated to run the assay. Corticosterone concentrations were measured using a commercial [^{125}I]-radioimmunoassay kit (MP Biomedicals) according to instructions. Samples were run in duplicate. The minimum detection limit of the assay was 7.7 ng ml^{-1} and the inter-replicate variability was 6.0%.

Gene expression analysis. For quantitative real-time PCR with reverse transcription (RT–PCR) experiments, mice were handled and sleep-deprived as described earlier. Hippocampal dissections were performed immediately after the deprivation, and alternated between SD and NSD animals. RNA preparation and real-time RT–PCR was performed as previously described³². In brief, hippocampi were flash-frozen in 500 μl RNeasy lysis buffer (Qiagen) and stored at -80°C . Aerosol Barrier ART tips, DEPC-treated sterile water (Ambion), and autoclaved 1.5-ml microcentrifuge tubes were used for all subsequent steps. Thawed hippocampi were homogenized on ice in 1 ml Trizol (Invitrogen) using a dounce homogenizer. Phenol–chloroform (300 μl) addition was followed by vigorous mixing and room-temperature incubation for 3 min. Samples were transferred to phase-lock gel tubes (Eppendorf) and centrifuged at 4°C and full speed for 15 min to extract RNA. The aqueous phase was transferred to new tubes, where 2 volumes ethanol, 1/10 volume 3 M sodium acetate, and 1 μl glycogen (10 mg ml^{-1}) were added. After gentle mixing, the samples were incubated at -20°C for 10 min. Centrifugation at 4°C and full speed was followed by

aspiration of the supernatant. Pellets were washed with 300 μl 70% ethanol, and then centrifuged at 4°C and full speed for 5 min. The supernatant was aspirated and samples were air-dried for approximately 10 min. After resuspension in 100 μl water, RNA was purified using the RNeasy system (Qiagen) according to the manufacturer's protocol. Residual DNA was removed by treatment with DNA-free (Ambion). RNA was then re-precipitated with 1 μl glycogen, 50 μl ammonium acetate and 250 μl ethanol. The pellet was then rinsed twice with 300 μl 80% ethanol, and resuspended in 24 μl RNase-free water. RNA concentration and purity was quantified by NanoDrop spectrophotometry (Thermo Fisher Scientific).

Generation of cDNA was carried out using the RETROscript kit (Ambion). Each reaction was performed with 1 μg RNA in a total volume of 20 μl composed as follows: 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl_2 , 5 mM dithiothreitol (DTT), 500 μM each dNTP, 5 μM random decamers, 10 units RNase inhibitor and 100 units MMLV-RT. As controls, further reaction mixtures were generated that lacked either reverse transcriptase or template mRNA. Subsequent reactions were allowed to proceed at 44°C for 1 h, followed by heat-inactivation at 100°C for 10 min.

For quantitative real-time RT–PCR, reactions were prepared in 96-well optical reaction plates (ABI) with optical adhesive covers (ABI). Each well contained 11.4 μl cDNA, 1 μl of 5 μM primer mix solution and 12.4 μl Power SYBR Green PCR Master Mix (ABI). Three technical replicates were used. Reactions were carried out in the ABI Prism 7000 with an initial activation at 50°C for 2 min followed by incubation at 95°C for 15 min, and 40 subsequent cycles of 95°C for 15 s, 56°C for 30 s, and 72°C for 30 s. Primer sequences were as follows: *Pde4a*: forward 5'-CACCTTCCTGCTGCTTCTCT-3', reverse 5'-CTGCTGGAAGG GTCTCTGTC-3'. *Pde4b*: forward 5'-CAGATCAGGAACAGGTGT-3', reverse 5'-AGCGTGAGATGCTTGTGTG-3'. *Pde4d*: forward 5'-AGATGAGT CCGTCTGGCAAC-3', reverse 5'-CGGCCTTTCTCTTCTCTCT-3'. Data was normalized to *Act1*, *Hprt* and *Tuba4a* before calculation of differences. Housekeeping primers were the same as described previously³². Relative quantification of gene expression was performed according to ABI's User Bulletin no.2. Fold change was calculated from the ΔC_t values with corrections for standard curve data from each gene and housekeeping gene expression levels for each sample based on the relative standard curve method described in the Applied Biosystems manual. For each sample, the ΔC_t was calculated against the mean for that gene's sample set. Next, each of these ΔC_t values was corrected with the slope of the standard curve for the relevant primer set to account for any variation in primer amplification efficiency. The efficiency-corrected ΔC_t value was normalized to the similarly corrected ΔC_t from housekeeping genes for each sample to account for variability in mRNA input. The difference between corrected ΔC_t for each sample was then grouped according to the type (SD or NSD) and the average ΔC_t was calculated. Because C_t values are on a logarithmic scale, fold change is equal to two raised to the difference between experimental and control ΔC_t values. Because corrections were made for primer efficiency, we have presented the data as fold change. The data presented is the calculated mean for the biological replicates with n being equal to the number of biological replicates (that is, the number of mice examined).

Statistics. For LTP experiments, the maintenance of LTP was analysed using repeated measures one-way analysis of variance (ANOVA) tests with Tukey post-hoc tests, with group and time as factors, and the percentage baseline fEPSP slope during the last 20 min of the recording as the variable. For experiments using rolipram treatment, repeated measures two-way ANOVA tests were used, with group and treatment as factors. Analysis of LTP induction was carried out using t -tests, except in the case of experiments using rolipram, in which two-way ANOVAs were used, with group and treatment as factors. To evaluate differences in input–output characteristics, a t -test was performed comparing the average linear regression slopes for sleep-deprived and control mice. For paired-pulse facilitation, a repeated measures two-way ANOVA with Tukey post-hoc tests was used with group and inter-stimulus interval as factors and the ratio of the initial fEPSP slopes elicited by the two stimuli as the variable. For whole-cell patch-clamp recordings, NMDA-receptor I – V curves were analysed using a two-way ANOVA with voltage and sleep deprivation as factors. NMDA-receptor input–output curves were compared using a two-way ANOVA with stimulation intensity and sleep deprivation as factors. NMDA receptor/AMPA receptor ratios were compared between sleep-deprived and non-sleep-deprived animals using a t -test. For cAMP assays, a non-parametric Kruskal–Wallis test was used because data across treatment groups was non-normal. An overall effect of group was found ($H(5, N = 108) = 42.04, P < 0.0001$), and P values reported in the text are from post-hoc Mann–Whitney U tests used to make the following planned comparisons: NSD control versus SD control, NSD + FSK versus SD + FSK, and NSD + FSK/IBMX versus SD + FSK/IBMX. For PDE activity assays and PDE4 western blots, t -tests were used to compare values between SD and NSD animals, using one-tailed P values because of our prediction that PDE levels

and activity would be increased by sleep deprivation. For pCREB immunohistochemistry, *t*-tests were used to compare optical densities in each hippocampal region, using one-tailed *P* values because of our initial prediction that CREB phosphorylation would be reduced by sleep deprivation. For quantitative real-time RT-PCR experiments, *t*-tests were used to compare fold change values for each gene between sleep-deprived and control mice, and one-tailed *P* values are reported because of our initial prediction that PDE expression would be increased by sleep deprivation. Two-tailed *P* values are reported for RT-PCR experiments examining the effects of recovery sleep. For fear conditioning experiments, a two-way repeated measures ANOVA with Tukey post-hoc tests was used, with group (SD versus NSD) and drug treatment (rolipram versus vehicle) as factors, and context (trained, altered) as the repeated measure, and percentage freezing as the variable. Effects on context-specific freezing were measured using a two-way ANOVA with Tukey post-hoc tests, with group and drug treatment as factors, and the difference in percentage freezing between the trained and altered contexts as the variable. All statistics were carried out using Statistica 7 (Statsoft, Inc.). In all figures in the main text, asterisk represents a comparison with $P < 0.05$, and all error bars indicate \pm s.e.m.

31. Ledoux, L., Sastre, J. P., Buda, C., Luppi, P. H. & Jouvet, M. Alterations in c-fos expression after different experimental procedures of sleep deprivation in the cat. *Brain Res.* **735**, 108–118 (1996).
32. Vecsey, C. G. *et al.* Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB:CBP-dependent transcriptional activation. *J. Neurosci.* **27**, 6128–6140 (2007).
33. Fanselow, M. S. Conditioned and unconditional components of post-shock freezing. *Pavlov. J. Biol. Sci.* **15**, 177–182 (1980).
34. Matthies, H. *et al.* Design of a multiple slice interface chamber and application for resolving the temporal pattern of CREB phosphorylation in hippocampal long-term potentiation. *J. Neurosci. Methods* **78**, 173–179 (1997).
35. Malleret, G. *et al.* Inducible and reversible enhancement of learning, memory, and long-term potentiation by genetic inhibition of calcineurin. *Cell* **104**, 675–686 (2001).
36. Abel, T. *et al.* Genetic demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-term memory. *Cell* **88**, 615–626 (1997).
37. Huang, Y. Y. & Kandel, E. R. Recruitment of long-lasting and protein kinase A-dependent long-term potentiation in the CA1 region of hippocampus requires repeated tetanization. *Learn. Mem.* **1**, 74–82 (1994).
38. Woo, N. H., Duffy, S. N., Abel, T. & Nguyen, P. V. Temporal spacing of synaptic stimulation critically modulates the dependence of LTP on cyclic AMP-dependent protein kinase. *Hippocampus* **13**, 293–300 (2003).
39. Nguyen, P. V., Abel, T. & Kandel, E. R. Requirement of a critical period of transcription for induction of a late phase of LTP. *Science* **265**, 1104–1107 (1994).
40. Nguyen, P. V. & Kandel, E. R. Brief theta-burst stimulation induces a transcription-dependent late phase of LTP requiring cAMP in area CA1 of the mouse hippocampus. *Learn. Mem.* **4**, 230–243 (1997).
41. Daly, J. W., Padgett, W. & Seamon, K. B. Activation of cyclic AMP-generating systems in brain membranes and slices by the diterpene forskolin: augmentation of receptor-mediated responses. *J. Neurochem.* **38**, 532–544 (1982).
42. Beavo, J. A. *et al.* Effects of phosphodiesterase inhibitors on cyclic AMP levels and on lipolysis. *Ann. NY Acad. Sci.* **185**, 129–136 (1971).
43. Conti, M. & Beavo, J. Biochemistry and physiology of cyclic nucleotide phosphodiesterases: essential components in cyclic nucleotide signaling. *Annu. Rev. Biochem.* **76**, 481–511 (2007).
44. Wu, L. J. *et al.* Neurabin contributes to hippocampal long-term potentiation and contextual fear memory. *PLoS One* **3**, e1407 (2008).
45. Marchmont, R. J. & Houslay, M. D. Insulin trigger, cyclic AMP-dependent activation and phosphorylation of a plasma membrane cyclic AMP phosphodiesterase. *Nature* **286**, 904–906 (1980).
46. Lobban, M., Shakur, Y., Beattie, J. & Houslay, M. D. Identification of two splice variant forms of type-IVB cyclic AMP phosphodiesterase, DPD (rPDE-IVB1) and PDE-4 (rPDE-IVB2) in brain: selective localization in membrane and cytosolic compartments and differential expression in various brain regions. *Biochem. J.* **304**, 399–406 (1994).
47. Huston, E. *et al.* The cAMP-specific phosphodiesterase PDE4A5 is cleaved downstream of its SH3 interaction domain by caspase-3. Consequences for altered intracellular distribution. *J. Biol. Chem.* **275**, 28063–28074 (2000).
48. Lynch, M. J. *et al.* RNA silencing identifies PDE4D5 as the functionally relevant cAMP phosphodiesterase interacting with beta arrestin to control the protein kinase A/AKAP79-mediated switching of the β 2-adrenergic receptor to activation of ERK in HEK293B2 cells. *J. Biol. Chem.* **280**, 33178–33189 (2005).

LETTERS

The postsynaptic function of type II cochlear afferents

Catherine Weisz¹, Elisabeth Glowatzki^{1,2} & Paul Fuchs^{1,2}

The mammalian cochlea is innervated by two classes of sensory neurons. Type I neurons make up 90–95% of the cochlear nerve and contact single inner hair cells to provide acoustic analysis as we know it. In contrast, the far less numerous type II neurons arborize extensively among outer hair cells (OHCs)^{1,2} and supporting cells^{3,4}. Their scarcity and smaller calibre axons have made them the subject of much speculation, but little experimental progress for the past 50 years. Here we record from type II fibres near their terminal arbors under OHCs to show that they receive excitatory glutamatergic synaptic input. The type II peripheral arbor conducts action potentials, but the small and infrequent glutamatergic excitation indicates a requirement for strong acoustic stimulation. Furthermore, we show that type II neurons are excited by ATP. Exogenous ATP depolarized type II neurons, both directly and by evoking glutamatergic synaptic input⁵. These results prove that type II neurons function as cochlear afferents, and can be modulated by ATP. The lesser magnitude of synaptic drive dictates a fundamentally different role in auditory signalling from that of type I afferents.

The organ of Corti was dissected from the apical turn of postnatal rat cochleas (postnatal days (P)5–19) and secured in a recording chamber. A few OHCs were removed by aspiration to reveal nerve fibres (~2 µm diameter) running along the cochlear spiral (Fig. 1a–c). Gigaohm-seal ruptured-patch recordings were performed from the fibres (Fig. 1a–c). AlexaFluor 488 hydrazide was included in the recording pipette for subsequent visualization by post-hoc immunolabelling (Fig. 1d). Tracings from two fills (Fig. 1e) show the minimally branched terminal field several hundred micrometres basal to a marked right-angle turn in the fibre towards the tunnel of Corti. These fills (in six fibres) revealed spiral processes 100–325-µm long that terminated among the OHCs, and variable filling of the radial, central-going process (Supplementary Table 1), in one case to its soma in the spiral ganglion (Fig. 1e). These morphological features accord with type II afferent innervation patterns^{2,6}.

Voltage-gated currents were elicited with a series of 10 mV steps from –110 to +30 mV (Fig. 1f). Positive to –60 mV, transient, tetrodotoxin-sensitive inward currents were evoked (Fig. 1f, inset). On the basis of their all-or-none appearance, these are probably ‘action currents’, arising in distant, unclamped membrane. Positive to –50 mV sustained outward currents were also evoked. These were not characterized further, except to note that they were reduced when caesium replaced potassium in the recording pipette.

In current-clamp recording, action potentials were evoked when type II fibres were depolarized with injected current (Fig. 1g). Small excitatory postsynaptic potentials (EPSPs) were observed that averaged 3.8 ± 2.0 mV in amplitude ($n = 1,709$ EPSPs, $n = 8$ fibres).

Under voltage-clamp, excitatory postsynaptic currents (EPSCs) (Fig. 2a and inset) occurred several times per minute under resting

conditions. When external potassium was increased from the normal 5.8 to 15 or 40 mM to depolarize presynaptic sources, EPSC frequency increased (Fig. 2a and Supplementary Table 2). Amplitude histograms typically peaked near 18 pA (holding potential –90 mV) and were slightly skewed towards larger amplitudes (Fig. 2b), reaching a maximum of ~100 pA in some fibres. The mean amplitude value from 30 fibres was 28.3 ± 8.3 pA.

Synaptic currents recorded at different holding potentials were averaged (Fig. 2c) to provide an *I–V* relation that reversed at 0 mV

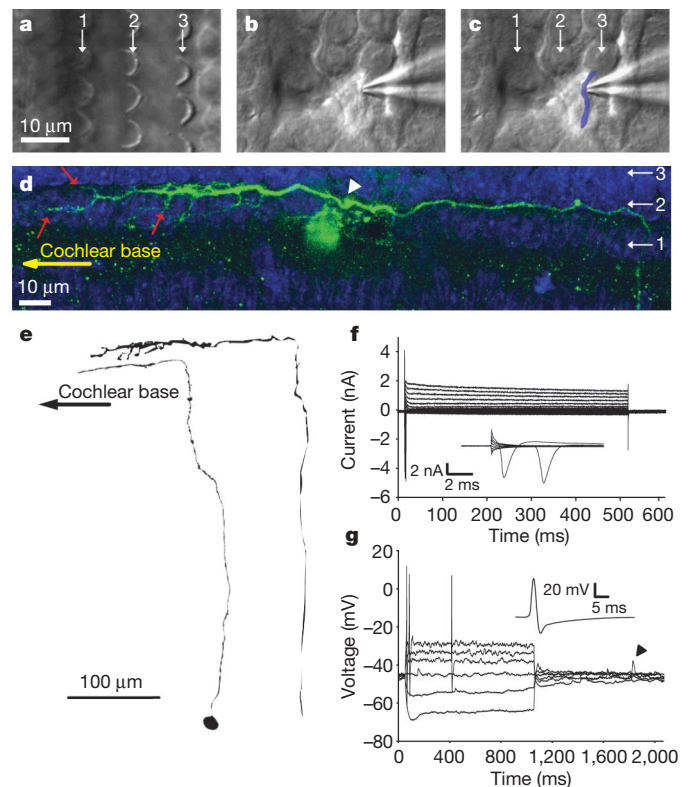


Figure 1 | Recording from type II terminal arbors. **a**, OHC cilia (white arrows). **b**, **c**, Pipette attached to type II fibre below OHCs. **d**, Confocal projection of dye-filled fibre (green). OHC nuclei (blue, 4',6-diamidino-2-phenylindole (DAPI)) visible in rows 1–3, red arrows indicate fibre branches towards OHCs. Recording site (white arrowhead) near dye artefact ‘cloud’. **e**, Drawings of fill from **d** (P6, OHC row 2) and another fibre (P5, OHC row 1). **f**, Currents evoked by 10 mV steps from –80 mV. Inset: selected inward currents, expanded. **g**, Current-clamp-evoked action potentials (threshold: -32.1 ± 10.0 mV). Resting potential -56.9 ± 10.2 mV ($n = 10$). Spontaneous action potential (inset) and small EPSPs (arrowhead). P5–P9 rats.

¹The Department of Neuroscience, ²The Department of Otolaryngology-Head and Neck Surgery, The Center for Hearing and Balance and the Center for Sensory Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

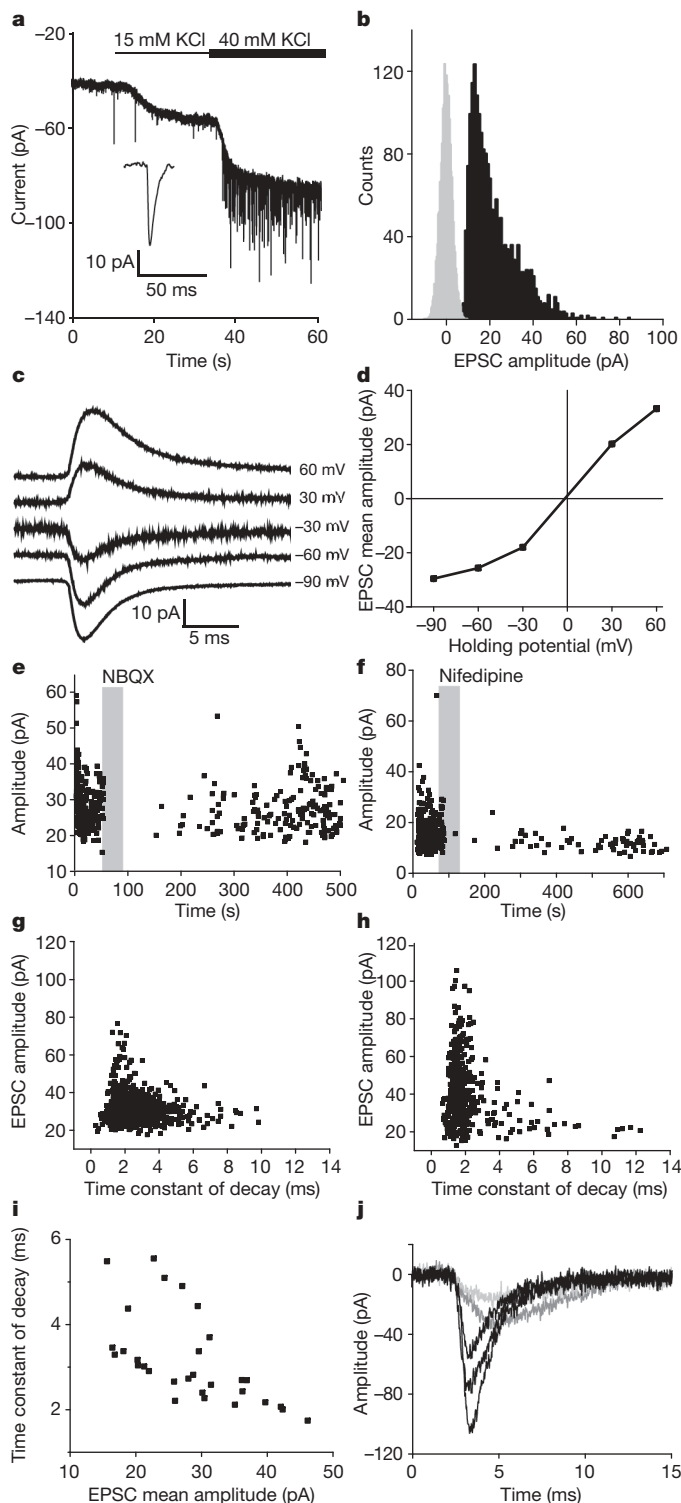


Figure 2 | EPSCs in type II fibres. **a**, Increased extracellular potassium-evoked EPSCs. Inset: EPSC waveform. **b**, Representative EPSC amplitude distribution (scaled noise in grey). **c**, **d**, Average EPSCs (**c**) and resulting I - V relation (**d**). **e**, EPSC diary plot showing reversible block by NBQX (10 μ M). **f**, EPSC diary plot showing reversible block by nifedipine (50 μ M). **g**, **h**, Amplitude versus decay time constant for EPSCs from two fibres. **i**, Mean decay time constant versus mean EPSC amplitude from 30 fibres. Linear regression fit ($F_{1,29} = 16.43$, $P = 0.004$; $r^2 = 0.37$). **j**, Exemplar EPSC waveforms (fibre in **h**). P5–P9 rats.

($n = 6$ fibres) (Fig. 2d). The AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)-type glutamate receptor antagonist NBQX reversibly blocked the EPSCs ($n = 7$ fibres) (Fig. 2e). Synaptic

currents in type II fibres were essentially eliminated by nifedipine ($n = 4$) (Fig. 2f) that blocks voltage-gated $\text{Ca}_v1.3$ calcium channels in OHCs⁷. Thus, EPSCs recorded from type II fibres are mediated by AMPA-type glutamatergic receptors, and are presumed to reflect transmitter release from OHCs.

Analysis of 12,043 EPSCs in 30 fibres from all rows gave average rise times of 1.1 ± 0.3 ms, and decay time constants of 3.2 ± 1.1 ms at room temperature (Supplementary Tables 2 and 3). Because each type II recording site might be anywhere within one length constant of presumptive inputs (Supplementary Table 1), synaptic waveforms could be altered by cable loss. There was a weak negative correlation between EPSC amplitude and time course in some fibres (Fig. 2g), whereas others were dominated by kinetically uniform EPSCs (Fig. 2h), as though cable effects varied between recordings. Among 30 fibres there was a significant correlation between the average EPSC amplitude and decay time (Fig. 2i), suggesting that recording sites were located at different distances from the synaptic inputs. Although type II EPSCs varied in amplitude and kinetics (Fig. 2j), these typically rose and fell smoothly, without the obvious inflections commonly found in all type I fibre recordings⁸.

It is known that OHCs are depolarized by ATP⁵ and that antibodies to P2X ATP receptors label type II fibres⁹. Thus, we next asked whether type II fibres were affected by ATP. The effect of ATP varied with postnatal age and will be described first for young (P5–P9) fibres. ATP applied from a nearby perfusion pipette led to frequent EPSCs (Fig. 3a). ATP evoked EPSCs at $2.4 \pm 2.1 \text{ s}^{-1}$ ($n = 10$ fibres), not significantly different from the effect of 40 mM K^+ that depolarizes OHCs to -35 mV ($n = 5$ cells) (Supplementary Table 2). ATP also evoked a prolonged inward current in type II fibres (Fig. 3a). The slow inward current seems to be a direct effect of ATP on the type II fibre because it preceded the EPSCs and was unchanged by the application of nifedipine that eliminated the ATP-evoked increase in EPSCs ($n = 4$, not shown). In current-clamp experiments, ATP depolarized type II fibres (31.3 ± 10.1 mV, $n = 4$) and induced EPSPs, although action potentials were rarely seen. However, this lack of action potentials may be an artefact of ‘wash-out’ because ATP reliably induced action potentials in type II fibres in extracellular, loose-patch recordings ($n = 7$) (Fig. 3b). This excitation was prevented by the P2X antagonist PPADS ($n = 4$, Fig. 3b). PPADS also eliminated ATP-evoked EPSCs observed during intracellular recording from type II fibres (not shown).

In newly hearing animals (P12–P13), type II fibres had voltage-gated sodium and potassium currents, and had EPSCs with average amplitudes and waveforms that were not different from those of younger fibres (not shown). ATP produced significantly less effect in voltage-clamp (Fig. 3c) and current-clamp (Fig. 3d). However, ATP was still able to evoke extracellularly recorded action potentials from four out of seven type II fibres at P13. Intracellular fibre recordings become exceedingly difficult with increasing animal age. Nonetheless, voltage-clamp and current-clamp recordings were achieved from several type II fibres at P17–P19, revealing EPSCs (Fig. 3e) and action currents as in younger fibres. In three out of six recordings, the application of ATP produced an effect on membrane current or membrane potential (Fig. 3f, g), although this was smaller than at younger ages.

Type II afferent function has been a mystery since the early twentieth century⁶. In contrast to the wealth of information on type I afferents, only one anatomically confirmed type II recording has been made *in vivo*, showing no spontaneous activity and no response to even very loud sound¹⁰. Voltage- and current-clamp recordings from type II neuronal somata *ex vivo* have begun to define voltage-gated conductances¹¹, and revealed slowly accommodating repetitive action potentials¹², but could not assess synaptic inputs. By recording near the terminal arbor, we show that type II afferents receive glutamatergic synaptic input and also are subject to purinergic activation.

The EPSCs in type II afferents were much less frequent, slower on average and significantly smaller than glutamatergic EPSCs in type I afferent boutons on inner hair cells (IHCs)^{8,13}. Cable loss probably

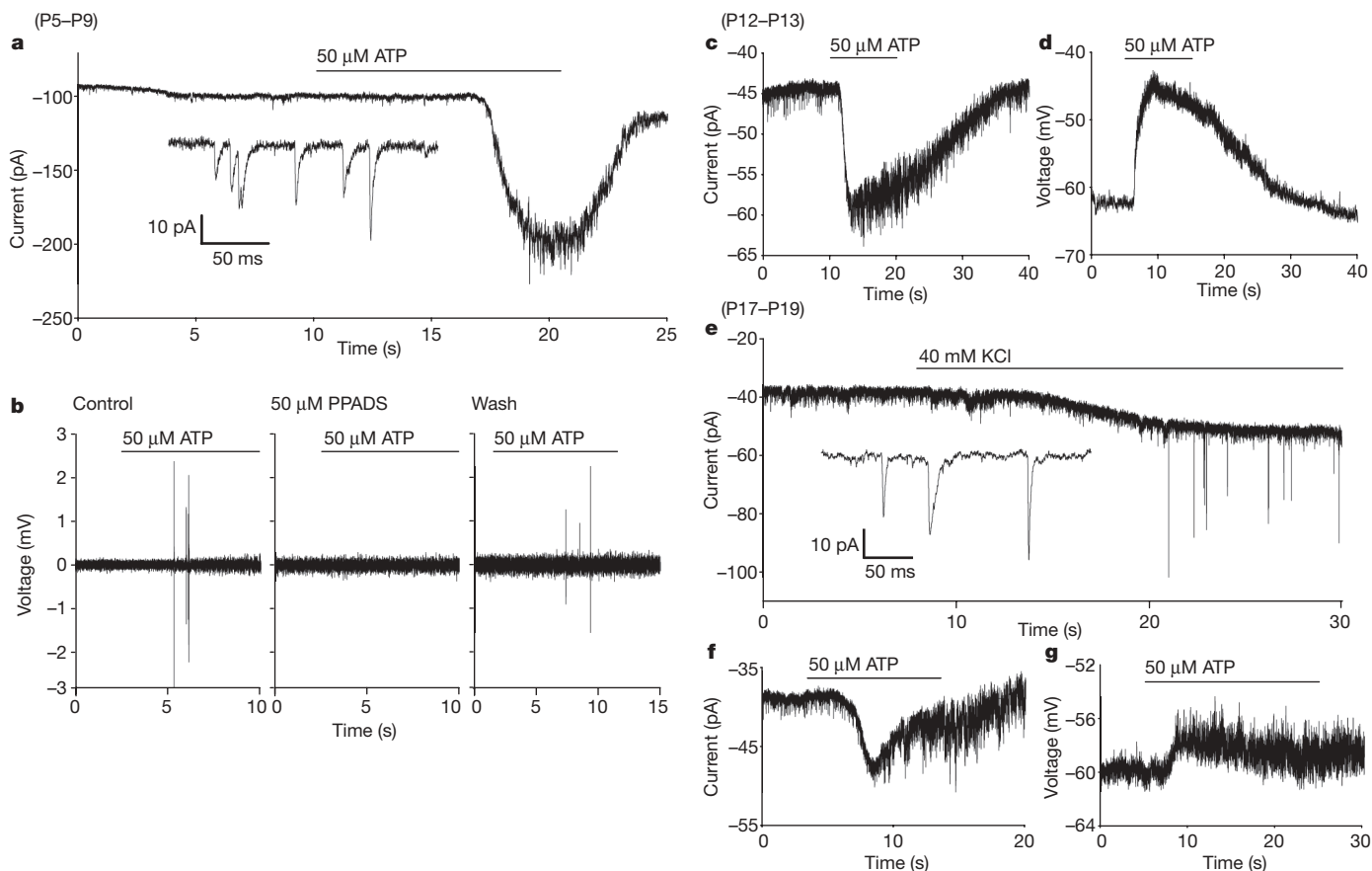


Figure 3 | ATP stimulates type II fibres. **a**, ATP-evoked inward currents in postnatal (P5–P9) fibres (142.7 ± 73.6 pA, $n = 6$) and increased EPSC frequency. Inset: ATP (1 μM) induced EPSCs in another cell. **b**, PPADS reversibly blocked ATP-induced repetitive action potentials (loose-patch extracellular record). **c**, ATP-induced inward current (29.9 ± 17.4 pA, $n = 5$)

and EPSCs in P12–P13 fibres. **d**, ATP depolarized P12–P13 fibres (12.6 ± 4.8 mV, $n = 4$). **e**, 40 mM extracellular potassium induced EPSCs in a P18 fibre. Inset: expanded waveforms, same cell. **f**, ATP evoked inward currents in three of six P17–P19 fibres (10.1 ± 3.6 pA, $n = 3$). **g**, ATP depolarized P17–P19 fibres (3.1 ± 0.2 mV, $n = 2$).

accounts for some differences in waveform, although other reports suggest the possibility of fundamental differences between type II and type I glutamatergic mechanisms^{14,15}. What is more certain is that presynaptic mechanisms differ substantially for type I and type II afferents. Considering that there are more than a dozen presynaptic OHCs, EPSCs were remarkably infrequent in type II afferents compared to much higher release rates from a single IHC ribbon synapse onto a type I afferent⁸ (Supplementary Table 2). This indicates that overall synaptic transfer is relatively poor from OHCs, consistent with reports of smaller voltage-gated calcium currents in OHCs compared with IHCs¹⁶. However, OHCs are capable of vesicular fusion¹⁷, and support it with the same dihydropyridine-sensitive calcium channels ($\text{Ca}_v1.3$) as in IHCs⁷. Furthermore, OHCs do have synaptic ribbons, although fewer than IHCs, not at every afferent contact, and their prevalence may differ between species^{18,19}. Type II afferents also contact Deiters and Hensen cells, but these are unlikely to contribute because they possess no presynaptic ultrastructure, and are described as postsynaptic to type II afferents^{20,21}.

Although type II synaptic organization remains incompletely resolved, it is clear from the present data that OHCs release vesicles infrequently, and the postsynaptic effect is relatively small. The dominant mode of EPSC amplitude distributions may correspond to the release of a single vesicle, in contrast to multivesicular release at the IHC afferent synapse^{8,22}. In keeping with that suggestion, ‘multiphasic’ EPSCs were not observed in the type II recordings, promising future insights into variation among hair cell release mechanisms. Whatever the underlying mechanism, the net result is that synaptic release from OHCs produced only modest depolarization of the type II fibre, and seems unlikely on its own to signal moderate acoustic

stimulation. Although both groups of afferents project together to the cochlear nucleus²³, the basalward spiral of type II afferents indicates an offset ($\sim 1/4$ octave) towards higher acoustic frequencies than those of associated type I afferents⁴. The present results require such higher frequency sound to be significantly louder, if type II activity were to be integrated centrally with that of neighbouring type I afferents.

Diverse roles have been suggested for the function of type II afferents, including measuring the ‘set-point’ of OHC electromotility¹¹, or regulating a local network of signalling between OHCs through reciprocal synapses¹⁴. The developmental fall in sensitivity to ATP observed here is consistent with the downregulation of some P2X receptors during development²⁴, and suggests a possible role in synaptic maturation like that reported for type I afferents²⁵. Finally, previous workers have proposed that type II afferents might mediate responses to loud, even painful sounds²⁶. Tissue damage triggers calcium waves in supporting cells that depend on release of ATP²⁷, and loud sound can increase ATP levels in cochlear fluids²⁸, reminiscent of the role of ATP in somatic pain²⁹. The participation of type II afferents would be strengthened by renewed sensitivity to ATP, and there is evidence that exposure to loud sound increases the expression of cochlear P2X receptors³⁰.

METHODS SUMMARY

The organ of Corti was dissected from the apical turn of postnatal (P5–P19) rat cochleas and secured in a recording chamber. Four-to-six OHCs were removed by aspiration to reveal nerve fibres (~ 2 μm diameter) running along the cochlear spiral. Gigaohm-seal ruptured-patch recordings were performed from the fibres using borosilicate glass electrodes with resistances of 6–10 MΩ. EPSCs were recorded in voltage-clamp at a membrane holding potential of -90 mV, unless otherwise noted. EPSPs were recorded in current-clamp with the fibre membrane potential held at -60 to -65 mV by direct current injection. Action

potentials were induced in current-clamp by depolarizing current injection in 10-pA steps. Action potentials were also recorded during loose-patch extracellular recordings performed with a seal resistance of 3–4 times the electrode resistance. Pharmacological compounds were applied with a gravity-fed, large bore application pipette placed near the recording site. AlexaFluor 488 hydrazide (10 μ M) was included in the recording pipette for diffusion into the fibre and subsequent visualization by post-hoc immunolabelling. The labelled fibre was imaged using a confocal microscope. Data are presented as the mean \pm s.d.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 May; accepted 7 September 2009.

- Perkins, R. E. & Morest, D. K. A study of cochlear innervation patterns in cats and rats with the Golgi method and Nomarski Optics. *J. Comp. Neurol.* **163**, 129–158 (1975).
- Berglund, A. M. & Ryugo, D. K. Hair cell innervation by spiral ganglion neurons in the mouse. *J. Comp. Neurol.* **255**, 560–570 (1987).
- Fechner, F. P., Nadol, J. J., Burgess, B. J. & Brown, M. C. Innervation of supporting cells in the apical turns of the guinea pig cochlea is from type II afferent fibers. *J. Comp. Neurol.* **429**, 289–298 (2001).
- Brown, M. C. Morphology of labeled afferent fibers in the guinea pig cochlea. *J. Comp. Neurol.* **260**, 591–604 (1987).
- Nakagawa, T., Akaïke, N., Kimitsuki, T., Komune, S. & Arima, T. ATP-induced current in isolated outer hair cells of guinea pig cochlea. *J. Neurophysiol.* **63**, 1068–1074 (1990).
- Lorente de No, R. The sensory endings in the cochlea. *Laryngoscope* **47**, 373–377 (1937).
- Michna, M. *et al.* Cav1.3 (α 1D) Ca^{2+} currents in neonatal outer hair cells of mice. *J. Physiol. (Lond.)* **553**, 747–758 (2003).
- Glowatzki, E. & Fuchs, P. A. Transmitter release at the hair cell ribbon synapse. *Nature Neurosci.* **5**, 147–154 (2002).
- Järleback, L. E., Housley, G. D. & Thorne, P. R. Immunohistochemical localization of adenosine 5'-triphosphate-gated ion channel P2X₂ receptor subunits in adult and developing rat cochlea. *J. Comp. Neurol.* **421**, 289–301 (2000).
- Robertson, D. Horseradish peroxidase injection of physiologically characterized afferent and efferent neurones in the guinea pig spiral ganglion. *Hear. Res.* **15**, 113–121 (1984).
- Jagger, D. J. & Housley, G. D. Membrane properties of type II spiral ganglion neurones identified in a neonatal rat cochlear slice. *J. Physiol. (Lond.)* **552**, 525–533 (2003).
- Reid, M. A., Flores-Otero, J. & Davis, R. L. Firing patterns of type II spiral ganglion neurons *in vitro*. *J. Neurosci.* **24**, 733–742 (2004).
- Goutman, J. D. & Glowatzki, E. Time course and calcium dependence of transmitter release at a single ribbon synapse. *Proc. Natl Acad. Sci. USA* **104**, 16341–16346 (2007).
- Thiers, F. A., Nadol, J. B. Jr & Liberman, M. C. Reciprocal synapses between outer hair cells and their afferent terminals: evidence for a local neural network in the mammalian cochlea. *J. Assoc. Res. Otolaryngol.* **9**, 477–489 (2008).
- Matsubara, A., Laake, J. H., Davanger, S., Usami, S. & Ottersen, O. P. Organization of AMPA receptor subunits at a glutamate synapse: a quantitative immunogold analysis of hair cell synapses in the rat organ of Corti. *J. Neurosci.* **16**, 4457–4467 (1996).
- Knirsch, M. *et al.* Persistence of $\text{Ca}_v1.3$ Ca^{2+} channels in mature outer hair cells supports outer hair cell afferent signaling. *J. Neurosci.* **27**, 6442–6451 (2007).
- Beurg, M. *et al.* Calcium- and otoferlin-dependent exocytosis by immature outer hair cells. *J. Neurosci.* **28**, 1798–1803 (2008).
- Dunn, R. A. & Morest, D. K. Receptor synapses without synaptic ribbons in the cochlea of the cat. *Proc. Natl Acad. Sci. USA* **72**, 3599–3603 (1975).
- Hashimoto, S. & Kimura, R. S. Computer-aided three-dimensional reconstruction and morphometry of the outer hair cells of the guinea pig cochlea. *Acta Otolaryngol. (Stockh.)* **105**, 64–74 (1988).
- Simmons, D. D. & Liberman, M. C. Afferent innervation of outer hair cells in adult cats: II. Electron microscopic analysis of fibers labeled with horseradish peroxidase. *J. Comp. Neurol.* **270**, 145–154 (1988).
- Burgess, B. J., Adams, J. C. & Nadol, J. B. Jr. Morphologic evidence for innervation of Deiters' and Hensen's cells in the guinea pig. *Hear. Res.* **108**, 74–82 (1997).
- Neef, A. *et al.* Probing the mechanism of exocytosis at the hair cell ribbon synapse. *J. Neurosci.* **27**, 12933–12944 (2007).
- Berglund, A. M. & Brown, M. C. Central trajectories of type II spiral ganglion cells from various cochlear regions in mice. *Hear. Res.* **75**, 121–130 (1994).
- Huang, L. C., Greenwood, D., Thorne, P. R. & Housley, G. D. Developmental regulation of neuron-specific P2X₃ receptor expression in the rat cochlea. *J. Comp. Neurol.* **484**, 133–143 (2005).
- Tritsch, N. X., Yi, E., Gale, J. E., Glowatzki, E. & Bergles, D. E. The origin of spontaneous activity in the developing auditory system. *Nature* **450**, 50–55 (2007).
- Brown, M. C., Liberman, M. C., Benson, T. E. & Ryugo, D. K. Brainstem branches from olivocochlear axons in cats and rodents. *J. Comp. Neurol.* **278**, 591–603 (1988).
- Gale, J. E., Piazza, V., Ciubotaru, C. D. & Mammano, F. A mechanism for sensing noise damage in the inner ear. *Curr. Biol.* **14**, 526–529 (2004).
- Muñoz, D. J., Kendrick, I. S., Rassam, M. & Thorne, P. R. Vesicular storage of adenosine triphosphate in the guinea-pig cochlear lateral wall and concentrations of ATP in the endolymph during sound exposure and hypoxia. *Acta Otolaryngol. (Stockh.)* **121**, 10–15 (2001).
- Burnstock, G. Purinergic receptors and pain. *Curr. Pharm. Des.* **15**, 1717–1735 (2009).
- Wang, J. C. *et al.* Noise induces up-regulation of P2X₂ receptor subunit of ATP-gated ion channels in the rat cochlea. *Neuroreport* **14**, 817–823 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Supported by NIDCD grants R01 DC000276 and R01 DC006476, T32 DC000023 and a grant from the Blaustein Pain Foundation of Johns Hopkins.

Author Contributions C.W. performed and analysed all experiments with further analysis from P.F. and E.G. C.W., E.G. and P.F. conceived the project, designed and discussed the experiments, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.F. (pfuchs1@jhmi.edu).

METHODS

Electrophysiological recordings from type II afferents. Sprague-Dawley rat pups (Charles River) of P5–P19 were anaesthetized with 0.45 mg per 10 g euthasol (Virbac AH, Inc.) or isoflurane (Vedco, Inc.) according to approved Johns Hopkins Institutional Animal Care and Use Committee (IACUC) guidelines. After ensuring deep anaesthesia with a foot pinch, the animals were decapitated, and the temporal bone containing auditory and vestibular peripheral organs was removed. The bone surrounding the cochlea was dissected away and the apical turn of the cochlear spiral was severed at the modiolus. The stria vascularis and tectorial membrane were removed. The entire cochlear turn including spiral ganglion and organ of Corti was mounted under an insect pin glued to a coverslip for electrophysiological experiments.

Standard gigohm seal whole-cell patch-clamp and loose-patch techniques were used to record from terminals of the type II afferent fibres under OHCs. Using differential interference contrast (DIC) optics, 4–6 OHCs were aspirated to expose the type II dendrites for gigohm-seal voltage-clamp recording. Extracellular solution was perfused through the recording chamber at a rate of 2–3 ml min⁻¹. The solution contained (in mM): 5.8 potassium chloride, 155 sodium chloride, 1.3 calcium chloride, 0.9 magnesium chloride, 0.7 sodium phosphate, 5.6 glucose, 10 HEPES, pH 7.4. Intracellular solution contained (in mM): 150 potassium chloride, 0.1 calcium chloride, 3.5 magnesium chloride, 5 EGTA, 5 HEPES, 2.5 sodium-ATP, pH 7.2. Chemicals were purchased from Sigma. In some experiments ATP was excluded from the intracellular solution or potassium was replaced with caesium chloride. 1-mm borosilicate glass pipettes (WPI) were Sylgard-coated (Corning) and fire-polished to resistances of 6–10 M Ω . All pharmacological compounds were from Tocris Bioscience unless otherwise stated. Doses used were the following: 50 μ M PPADS, 50 μ M nifedipine, 10 μ M NBQX, 1–50 μ M ATP. Pharmacological compounds were applied with a large bore pipette positioned near the recording site. Recordings were performed

using an Axopatch 200B amplifier (Axon Instruments), pClamp version 9.2 (Axon Instruments) and a Digidata 1322A board (Axon). Data were sampled at 50 kHz and low-pass filtered at 10 kHz. Input resistances averaged 550 M Ω , series resistances were on average 25.5 M Ω and were not corrected for. Membrane holding potential is given without liquid junction potential correction of -4 mV. Loose patch extracellular recordings were performed with a seal resistance 3–4 times the pipette resistance.

Data analysis. EPSCs were analysed with MiniAnalysis software (Synaptosoft) or Clampfit 9.2 (Axon Instruments). Figures were prepared in Origin 8.0 (Origin Labs) and Illustrator (Adobe). Statistical analysis was performed with JMP (SAS). All data given as mean \pm s.d.

Immunohistochemical enhancement of neuronal tracer. The fluorescent neuronal tracer AlexaFluor 488 hydrazide (Molecular Probes) was included in the recording pipette at 10 μ M. After recording, cochlear turns were fixed with 4% paraformaldehyde in PBS at pH 7.4 for 60 min. The tracer signal was enhanced by immunolabelling with anti-Alexa 488 antibody as follows: block 2 h at room temperature in 5% normal goat serum with 0.25% Triton X-100 in 60 mM PBS, incubate in primary antibody overnight at 4 °C (rabbit anti-Alexa 488, 1:1,000 in blocking buffer), rinse 3 \times 10 min in blocking buffer, incubate in secondary antibody 2 h at room temperature (goat anti-rabbit 1:2,000 in blocking buffer), rinse 1 \times 10 min in 60 mM PBS, incubate with DAPI nuclear stain (1:1,000 in dH₂O), rinse 2 \times 10 min 60 mM PBS, mount in FluorSave medium (Calbiochem).

Equipment and settings for digital images. Cochlear explants were viewed for electrophysiological experiments under a Axioscope microscope (Zeiss) using DIC with a \times 40 water immersion objective and a camera with contrast enhancement (DAGE). Images were collected using a Zeiss LSM 510 Meta confocal microscope. Labelled fibres were measured (LSM Image Browser). Tiled images were traced in Photoshop (Adobe).

LETTERS

Unexpected consequences of a sudden and massive transposon amplification on rice gene expression

Ken Naito^{1,2}, Feng Zhang^{1†}, Takuji Tsukiyama², Hiroki Saito², C. Nathan Hancock¹, Aaron O. Richardson¹, Yutaka Okumoto², Takatoshi Tanisaka² & Susan R. Wessler¹

High-copy-number transposable elements comprise the majority of eukaryotic genomes where they are major contributors to gene and genome evolution¹. However, it remains unclear how a host genome can survive a rapid burst of hundreds or thousands of insertions because such bursts are exceedingly rare in nature and therefore difficult to observe in real time². In a previous study we reported that in a few rice strains the DNA transposon *mPing* was increasing its copy number by ~40 per plant per generation³. Here we exploit the completely sequenced rice genome to determine 1,664 insertion sites using high-throughput sequencing of 24 individual rice plants and assess the impact of insertion on the expression of 710 genes by comparative microarray analysis. We find that the vast majority of transposable element insertions either upregulate or have no detectable effect on gene transcription. This modest impact reflects a surprising avoidance of exon insertions by *mPing* and a preference for insertion into 5' flanking sequences of genes. Furthermore, we document the generation of new regulatory networks by a subset of *mPing* insertions that render adjacent genes stress inducible. As such, this study provides evidence for models first proposed previously^{4–6} for the involvement of transposable elements and other repetitive sequences in genome restructuring and gene regulation.

In a previous study we discovered that rice strain EG4 has over 1,000 *mPing* elements whereas the sequenced rice genome (Nipponbare) has only 50^{3,7,8}. In addition, with almost ~40 new inserts per plant, even small populations of EG4 contain thousands of new insertions. Because the transposable element burst is occurring in a species with a complete genome sequence, virtually all insertion sites can be determined. To this end, we used 454 sequencing⁹ to identify *mPing* insertion sites in a small population (24 plants) of selfed progeny from a single parent. Sequences flanking *mPing* insertions were amplified by vectorette PCR¹⁰ using primers with barcodes to identify the plant of origin of new insertions (Supplementary Fig. 1). In all, 928 insertions were detected in more than two EG4 siblings (shared insertions) and 736 insertions were individual-specific (unshared; from 20–65 per plant). All 1,664 insertions were characterized using the rice annotation project database¹¹ (<http://rapdb.dna.affrc.go.jp/>) (Supplementary Tables 1 and 2). A control dataset of 1,664 randomly selected genomic sites was generated 1,000 times for comparison (see Methods).

The vast majority of insertions, whether shared or unshared, were into single copy sequences of the genome (1,521 of 1,664 = 91.4% compared with control, 898 of 1,664 = 53.9%). Superimposition of all insertion sites on the physical map of the 12 rice chromosomes shows that both shared and unshared sites are enriched in euchromatic regions of high gene density and largely absent from heterochromatic regions (Fig. 1a and Supplementary Fig. 2). A preference for insertion into genic regions was noted previously for other DNA

transposons^{12,13}. However, the availability of over 1,500 insertions that are near or in genes, of which almost half are *de novo*, afforded a unique opportunity to investigate whether a successful element like *mPing* has evolved target preferences that serve to mitigate the impact of a massive and rapid increase in copy number.

A closer look at the insertion sites in rice genes indicates that both shared and unshared insertions of *mPing* are significantly underrepresented in coding exons (9 of 928 = 1.0% of shared and 6 of 736 = 0.8% of unshared compared with 103.4 of 1,664 = 6.3% of the control; $P \ll 0.01$) (Fig. 1b and Supplementary Tables 3 and 4). A similar distribution for most (old) transposable elements in eukaryotic genes has been attributed to selection acting to filter detrimental exon insertions³, but lack of unshared inserts in exons suggests that *mPing* avoids exons. One could argue however that even *de novo* exon insertions may have been filtered by selection (for example, gametophyte- or dominant-lethality). This argument, however, is not supported by previous studies where 30–35% of *de novo* insertions of *Tos17*^{4,15} and *Ac/Ds*¹⁶ in rice were into exons (Supplementary Table 5).

In contrast with insertions into coding regions, both shared and unshared insertions are significantly overrepresented in the vicinity of genes compared to the control data set (Fig. 1c, d and Supplementary Tables 3 and 4). The most dramatic deviation is within –1 kilobase (kb) of the transcription start site (shared, 192 of 928 = 20.7%, and unshared, 114 of 736 = 15.5%, compared with control, 92.5 of 1,664 = 5.6%).

No single mechanism can account for the biased distribution of *mPing* inserts. For example, *mPing* may avoid exons because its 9 base pair (bp) target site preference is (A+T)-rich (GC content = 31%, Supplementary Fig. 3) whereas rice exons are, on average, (G+C)-rich (~55%)¹⁷. However, *mPing* does not avoid the (G+C)-rich 5' untranslated region and is enriched just upstream of the transcription start site. An understanding of the mechanisms underlying these preferences is beyond the scope of this study as they may be influenced by other factors such as chromatin structure^{18,19}, which, so far, has not been thoroughly characterized in rice.

Taken together the analysis of 1,664 *mPing* insertion sites indicates that the spectrum of shared and unshared insertions is virtually indistinguishable. Furthermore, the position of insertions suggests that the rapid amplification of successful transposable elements like *mPing* may have a more modest impact on the host than previously thought because of highly evolved targeting mechanisms that minimize deleterious effects on host gene expression.

To test this hypothesis, we used comparative microarray analysis to assess the impact of *mPing* amplification on the transcription of rice genes. We focused on the expression of 710 genes harbouring *mPing* inserts (Fig. 2a; within 5 kb upstream and downstream) that are present in all EG4 plants (shared inserts) (928 total shared

¹Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA. ²Division of Agronomy and Horticulture Science, Graduate School of Agriculture, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan. [†]Present address: Department of Genetics, Cell Biology and Development, University of Minnesota, Minnesota 55455, USA.

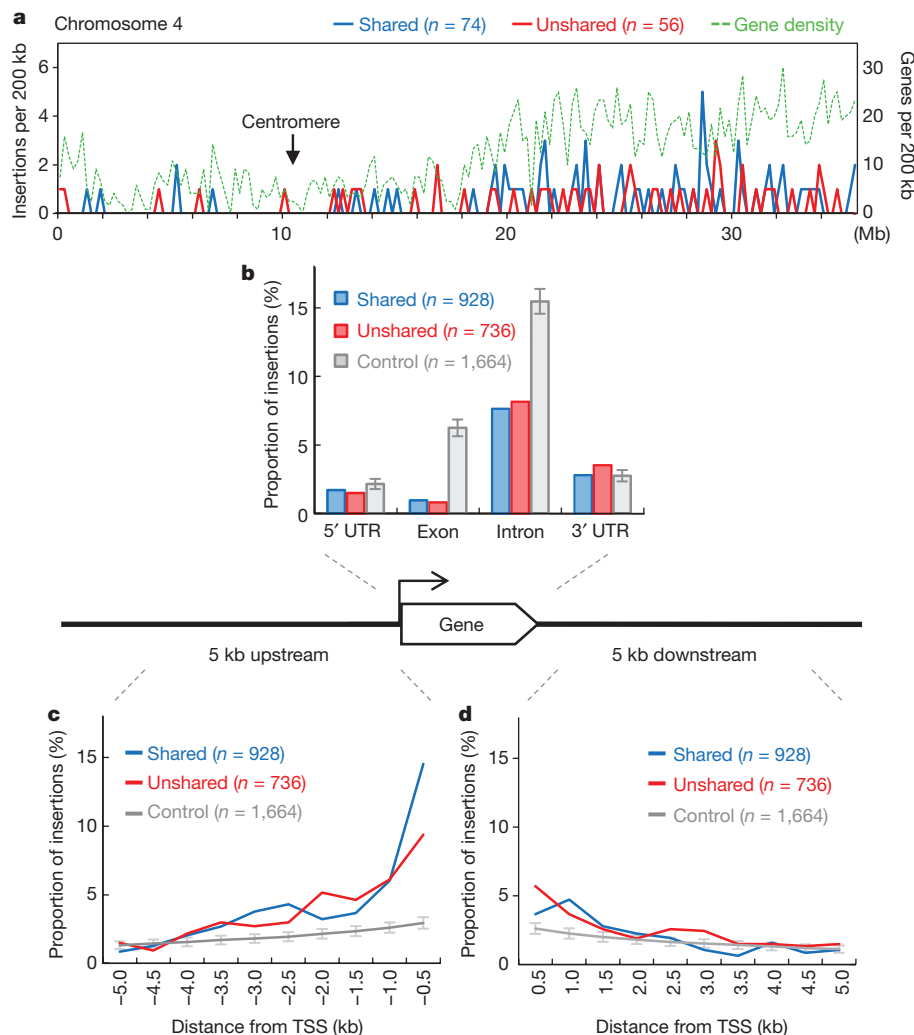


Figure 1 | Distribution of *mPing* insertions in strain EG4. **a**, Distribution along chromosome 4. The *x* axis indicates the distance along the chromosome and the *y* axes indicate the number of *mPing* insertions (left) and genes (right) per 200 kb. **b–d**, Proportion of all *mPing* insertions within

the coding region of annotated genes (**b**) and upstream (**c**) or downstream of annotated genes (**d**). TSS, transcription start site; TTS, transcription termination site; UTR, untranslated region. Mean \pm s.d., $n = 1,000$ (for 'control').

inserts – 218 intergenic shared inserts = 710; Supplementary Table 3). Unshared insertions were excluded from this analysis because most do not exist in the EG4 plants used for RNA isolation. To perform a comparative analysis of the 710 EG4 alleles (with *mPing*) to the Nipponbare alleles (without *mPing*, called NB alleles hereafter) RNA from seedlings of both strains grown under normal conditions was hybridized to 44K microarray chips containing 31,439 rice genes (see Methods)²⁰. Aside from the difference in *mPing* copy number, NB and EG4 are genetically very similar as EG4 is one of the strains of the cultivar Gimbozu, which is a great-grandparent of NB (see Supplementary Fig. 4 where the extreme *mPing* polymorphism is compared to the monomorphic pattern of the *Dasheng* retrotransposon²¹). As another measure of genetic similarity, only 7.0% of simple sequence repeat sites are polymorphic between EG4 and NB²².

Comparison of the expression of 31,439 EG4 and NB alleles showed that 82% are indistinguishable under normal growth conditions (Supplementary Table 6, 25,809 of 31,439). Of the remaining 18%, approximately half of the EG4 alleles are upregulated relative to the NB alleles whereas half are downregulated. The distribution of this 'background' of expression differences between EG4 and NB is reported as the control (see Fig. 2b, c).

Comparison of the distribution of gene expression differences between the 710 EG4 (*mPing*-containing) and NB (no *mPing*) alleles and controls reveals significant differences (Fig. 2b, c and Supplementary Tables 6 and 7). First, among the EG4 alleles that show

a difference in expression (156 of 710) there are significantly more upregulated EG4 alleles (111 of 156), largely due to alleles harbouring *mPing* within 1–5 kb upstream of the transcription start site ($P < 0.01$). Although there were proportionally more upregulated EG4 alleles with *mPing* in promoters (~ -1 kb), introns and downstream of the transcription termination site ($\sim +5$ kb), these were not significantly different from the control. Second, as expected, a significantly higher proportion of the 51 EG4 alleles with exon insertions were downregulated compared with the control ($P < 0.01$). Finally, microarray data were independently verified for 17 alleles using quantitative real-time PCR (qPCR) (Supplementary Figs 5 and 6).

The microarray data support the hypothesis that under normal growth conditions *mPing* insertions have a modest impact on host transcription, as the vast majority of EG4 alleles either have no impact or preferentially enhance transcription.

The enhancement of transcription by *mPing* motivated us to survey *mPing* sequence for regulatory motifs. Scanning *mPing* sequences with the plant *cis*-element database²³ (<http://www.dna.affrc.go.jp/PLACE/index.html>) identified 96 putative regulatory motifs; about one-third of which were stress responsive (Supplementary Table 8).

The involvement of a variety of stresses has been correlated with the activation of transposable elements ever since their discovery in a maize strain undergoing the breakage-fusion-bridge cycle²⁴. The *mPing* element is a deletion derivative of *Ping*, an autonomous

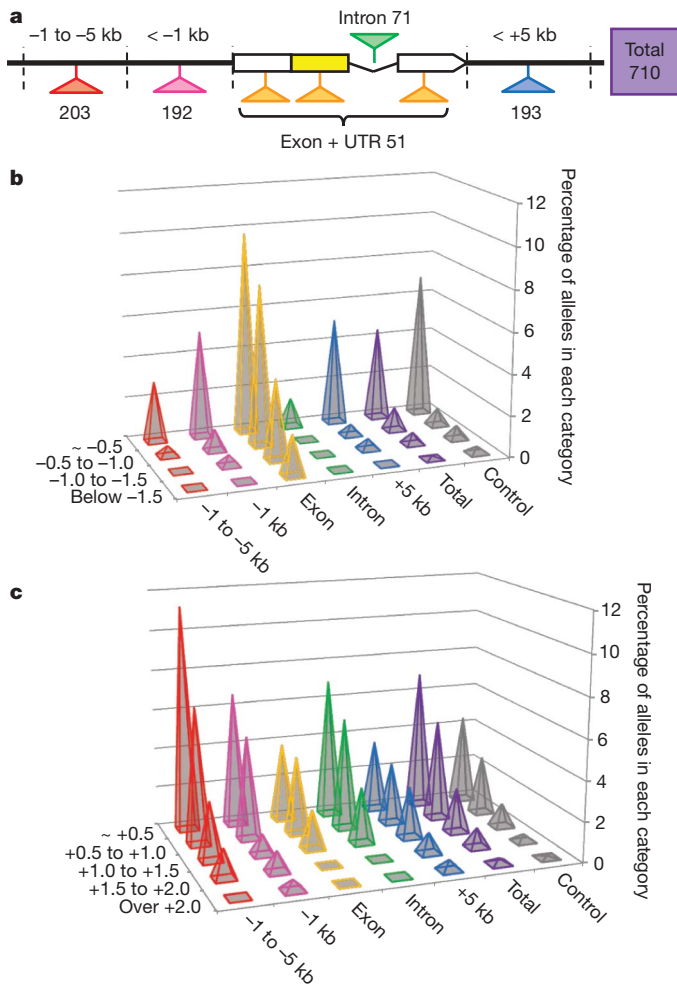


Figure 2 | Impact of *mPing*-containing alleles on host transcription.

a, Schematic of *mPing*-harbouring alleles from the shared insertions. Triangles represent *mPing* insertions that are colour-coded and include the number of insertions in the region shown. Yellow and white boxes represent coding and non-coding exons, respectively. **b**, **c**, Distribution of the degree of expression differences of *mPing*-harbouring alleles (in strain EG4) against Nipponbare alleles from microarray analyses among downregulated alleles (**b**) and upregulated alleles (**c**). The *x* axis is the degree of expression differences (\log_{10} value of fold change). The *y* axis is the proportion of all *mPing* alleles in the corresponding region.

element that codes for two proteins that are both required for transposition of *Ping* and *mPing* (Fig. 3a)²⁵. As such, *mPing* contains all of the sequences upstream of the basal promoter of ORF1. Whether these sequences are sufficient to confer stress-inducibility was tested by generating transgenic *Arabidopsis* plants containing this region fused to the GUS coding region. As shown in Fig. 3b, seedlings exposed to cold showed more intense GUS staining, which correlated with higher levels of GUS RNA in cold-stressed seedlings (Fig. 3c).

To investigate whether the presence of *mPing* conferred stress inducibility on nearby rice genes we began by performing qPCR on the products of 10 genes whose transcription was unaffected (in the microarray analysis) by the presence of *mPing* within 55 bp of the transcription start site. In addition, we exploited the availability of two landraces (A123, A157) that are very closely related to EG4 and have over 1,000 *mPing* inserts, but the spectrum of *mPing* insertions is dramatically different (Supplementary Fig. 4)³. As such, these strains allow us to compare, for example, the expression of a NB-type allele (no *mPing*) to an EG4-type allele (with *mPing*) where both are in strains with ~1,000 *mPing* insertions.

For each gene, we tested RNA isolated from the seedlings of four strains (NB, EG4, A123, A157) under conditions of cold, salt and dehydration stress (see Methods). For each allele there are landraces

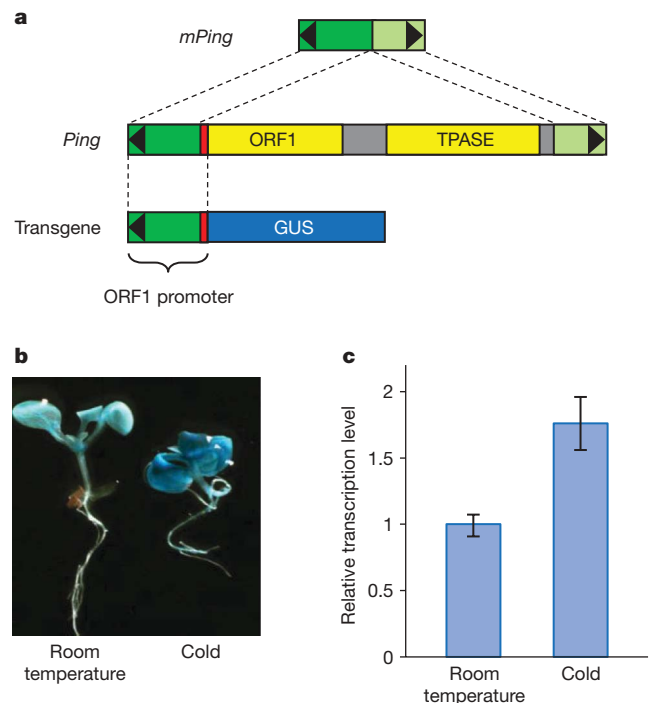


Figure 3 | Cold inducibility of a transgene in *Arabidopsis* containing the *Ping* ORF1 promoter fused to GUS. **a**, Derivation of *mPing* and a transgene promoter from the *Ping* element. Black triangles represent the terminal inverted repeat (TIR) and the red region represents the basal promoter of *Ping* ORF1. **b**, GUS staining of *Arabidopsis* transformants at room temperature and after cold treatment. **c**, Relative transcription levels of the GUS transgene in *Arabidopsis* seedlings. Mean \pm s.d., *n* = 3.

that have an NB-type allele. Of the ten loci tested, seven EG4-type alleles are inducible by cold and salt, but not by dehydration (Fig. 4 and Supplementary Fig. 7). Furthermore, sequencing of the NB and EG4 alleles of the three genes shown in Fig. 4 demonstrated that they differed only by the *mPing* insertion, except for a single nucleotide difference in one of the alleles (at -364 nucleotides in Os02g0582900).

To address the question of whether *mPing* is providing the promoter for nearby genes or whether it contains sequences that activate transcription from a distance, we analysed the RNA levels (at room temperature and after cold exposure) of genes where *mPing* was in introns (Supplementary Figs 8 and 9), ~2.5 kb upstream of the transcription start site (Supplementary Fig. 10) or downstream of the transcription termination site (Supplementary Fig. 11). We find that sequences within *mPing* act as an enhancer as the majority of EG4 alleles are cold-inducible regardless of the location of the insertion, while five of five control alleles (no *mPing* in any strain) showed virtually no difference in stress response (Supplementary Fig. 12). These data suggest that sequences within *mPing* provide new binding sites for transcription factors or other regulatory proteins.

In this study we have caught a transposon in the act of rapid amplification. Although conventional wisdom is that most transposable element insertions are neither beneficial nor neutral to the host, the results of our study demonstrate that populations can survive rapid and massive increases in transposable element copy number, even of transposable elements that prefer to insert into genic regions, because (successful) transposable elements have evolved target preferences that are largely neutral. Furthermore, we demonstrate that a large subset of the new alleles may actually benefit the host by creating potentially useful allelic variants and novel, stress-inducible regulatory networks. At least two stresses have been identified in this study (cold and salt), and one (drought) has been ruled out. It is possible that other subsets of *mPing* alleles respond to additional growth conditions. Taken together, *mPing* amplification can potentially create populations of rice (larger than the 24 rice plants in this study) with

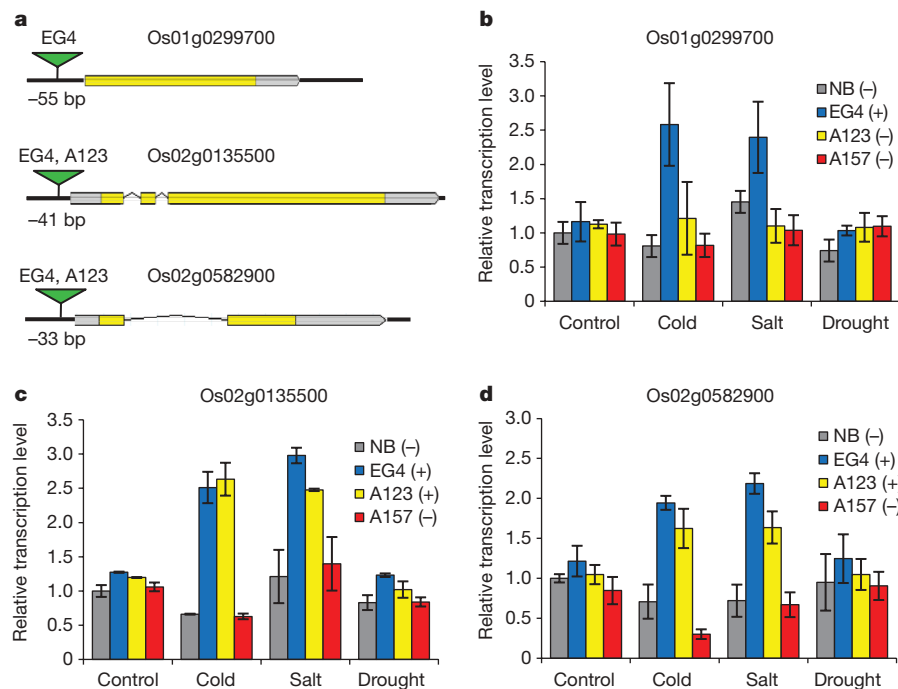


Figure 4 | Stress-induction of transcription of alleles harbouring *mPing*. **a**, Schematic of the gene structure of alleles with *mPing* in EG4 (top) and in strain A123. **b–d**, Relative transcription levels of selected alleles under stress conditions. The y axis is the amount of transcript in each sample relative to the

tens, perhaps hundreds of thousands of new alleles. Of note is that many of these alleles will be dominant and, as such, one allele, or a subset of alleles, could quickly alter one or more quantitative trait. This is consistent with McClintock's view of transposable elements as agents of evolutionary change^{4,26}. For rice and other selfing plants, transposable element bursts may be one of the critical solutions to rapidly generate genetic diversity in the face of an ever-changing environment. Furthermore, because evidence for the rapid and massive amplification of miniature inverted-repeat transposable elements has been found in virtually all sequenced eukaryotic genomes²⁷ (and even some prokaryotes²⁸), features of transposable element amplification documented for *mPing* are likely to be widespread in nature.

METHODS SUMMARY

Genomic DNA extracted from 24 EG4 plants was digested with BfaI and ligated to vectorette adapters¹⁰. Sequences flanking *mPing* insertions were amplified by nested PCR with *mPing*-specific primers and adapter primers with distinct four-base barcodes, pooled and subjected to pyrosequencing (454 Life Sciences). All sequences were assigned into distinct groups based on the DNA barcode. Shared and unshared *mPing* insertions were identified between barcoded groups by comparing non-redundant insertion sites. For the rice 44K microarray (Agilent Technologies) analysis, RNA was extracted from 10 seedlings each (seven-day-old) of EG4 and Nipponbare. EG4 versus NB alleles selected as up- or downregulated were determined according to the manufacturer's protocol. The *Ping* ORF1 promoter was cloned into the pDONR vector (Invitrogen), transferred to pMDC162 (ref. 29) by recombination, and transformed into *Arabidopsis thaliana* ecotype Columbia by the floral dip method³⁰. Whether EG4 landraces (A123 and A157) contain EG4 alleles (with *mPing*) or NB alleles (no *mPing*) was determined by PCR using primers flanking *mPing* inserts. For analysis of stress inducibility, 10-day-old plants were exposed to 4 °C for 2 h (cold), 150 mM NaCl for 24 h (salt), or deprived of water (dehydration) for 5 days. Microarray data can be found at GEO accession number GSE15021.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 July; accepted 2 September 2009.

1. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9**, 397–405 (2008).

RNA in NB, which is given a value of 1. RNA from two landraces (A123 and A157) where *mPing* has amplified to over 1,000 copies was also quantified. A '+' indicates that this strain shares the *mPing* allele with EG4 and a '-' indicates that the strain has the NB allele (no *mPing*). Mean \pm s.d., $n = 3$.

2. Feschotte, C., Jiang, N. & Wessler, S. R. Plant transposable elements: where genetics meet genomics. *Nature Rev. Genet.* **3**, 329–341 (2002).
3. Naito, K. et al. Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl Acad. Sci. USA* **103**, 17620–17625 (2006).
4. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
5. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).
6. Britten, R. J. & Davidson, E. H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46**, 111–138 (1971).
7. Jiang, N. et al. Active DNA transposon family in rice. *Nature* **421**, 163–167 (2003).
8. Kikuchi, K., Terauchi, K., Wada, M. & Hirano, H. Y. The plant MITE *mPing* is mobilized in anther culture. *Nature* **421**, 167–170 (2003).
9. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
10. Arnold, C. & Hodgson, I. J. Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl.* **1**, 39–42 (1991).
11. Itoh, T. et al. Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.* **17**, 175–183 (2007).
12. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* **140**, 315–324 (1995).
13. Liao, G. C., Rehm, E. J. & Rubin, G. M. Insertion site preferences of the *P* transposable element in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 3347–3351 (2000).
14. Miyao, A. et al. Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15**, 1771–1780 (2003).
15. Piffanelli, P. et al. Large-scale characterization of *Tos17* insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* **65**, 587–601 (2007).
16. Kolesnik, T. et al. Establishing an efficient Ac/Ds tagging system in rice: Large-scale analysis of Ds flanking sequences. *Plant J.* **37**, 301–314 (2004).
17. Yu, J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
18. Mavrich, T. N. et al. Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
19. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
20. Shiono, M. et al. WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell* **19**, 2064–2076 (2007).
21. Jiang, N. et al. *Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* **161**, 1293–1305 (2002).
22. Monden, Y. et al. High potential of a transposon *mPing* as a marker system in *japonica* \times *japonica* cross in rice. *DNA Res.* **16**, 131–140 (2009).

23. Higo, K., Ugawa, Y., Iwamoto, M. & Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.* **27**, 297–300 (1999).
24. McClintock, B. The relation of homozygous deficiencies to mutations and allelic series in maize. *Genetics* **29**, 478–502 (1944).
25. Yang, G., Zhang, F., Hancock, C. N. & Wessler, S. R. Transposition of a rice miniature inverted-repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **104**, 10962–10967 (2007).
26. Capy, P., Gasperi, G., Biemont, C. & Bazin, C. Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**, 101–106 (2000).
27. Feschotte, C., Zhang, X. & Wessler, S. R. in *Mobile DNA II* Ch. 50 (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 1147–1158 (ASM Press, 2002).
28. Delihas, N. Small mobile sequences in bacteria display diverse structure/function motifs. *Mol. Microbiol.* **67**, 475–481 (2007).
29. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes *in planta*. *Plant Physiol.* **133**, 462–469 (2003).
30. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Rice Genome Resource Center for the use of the rice microarray system and Y. Nagamura and R. Motoyama for technical support; the GenBank project of the National Institute of Agrobiological Science in Japan for providing seeds of Gimbozu landraces (A123 and A157); and X. Zhang and C. Feschotte for critical discussions and reading of the manuscript. S.R.W. is funded by a NSF Plant Genome grant and the University of Georgia Research Foundation and T.Tanisaka by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Contributions K.N. and F.Z. performed 454 sequencing and analysed the data. A.O.R. provided statistical analyses. T.Tsukiyama and Y.O. performed microarray, and K.N. and H.S. analysed the data. C.N.H. performed *Arabidopsis* transformation. K.N. performed stress treatment and real-time PCR. K.N., F.Z., T.Tanisaka and S.R.W. contributed the experimental design and wrote the paper.

Author Information Sequence data has been submitted to GEO under accession number GSE15021. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.R.W. (sue@plantbio.uga.edu) or T.Tanisaka (tanisaka@kais.kyoto-u.ac.jp).

METHODS

Plants. Rice plants of Nipponbare, EG4 and its landraces A123 and A157 were grown in plastic pots in an environment-controlled greenhouse with a photoperiod of 14 h light (25–28 °C). DNA and RNA were extracted 7 days after germination. Before RNA extraction, plants were exposed to cold temperature (4 °C for 2 h), high-salinity (150 mM NaCl for 24 h), or dehydration (deprived of water supply for 5 days) stress.

Sequencing and determination of *mPing* insertion sites. To determine *mPing* insertion sites in 24 individual sibling EG4 plants and their parent, flanking DNA fragments of *mPing* insertions were amplified using vectorette PCR¹⁰. Each DNA sample (~500 ng) was first digested overnight with BfaI at 37 °C, then ligated to linkers and amplified by nested PCR. To sequence all of the samples in a single sequencing run, the primers used in the second round of PCR contained unique 4-bp barcodes at their 5' end. PCR products were column-purified with a PCR purification kit (Qiagen), pooled and subjected to pyrosequencing (454 Life Science, the University of Oklahoma).

All reads generated from pyrosequencing were processed as follows by a Python script. All sequences were first filtered to remove the sequences without discernable 5' *mPing* sequences and then assigned into distinct groups based on their DNA barcode. Within each group sequences flanking *mPing* were extracted and those larger than 25 bp were searched against the rice genome database (RAP-DB) using BLAST with default parameters. Information on the best hits from the BLAST search was parsed to determine the location of *mPing* insertion sites in the rice genome. Redundant *mPing* insertions, defined as those that hit the same position in the genome, were removed from each group. Shared and unshared (unique) *mPing* insertions were identified between barcode groups by comparing non-redundant insertion sites.

For the control data set, 1,664 genome coordinates across the 12 chromosomes were randomly generated, and information on the surrounding sequence was parsed from the rep.gff file from RAP Annotation (release 2) (http://rapdb.dna.affrc.go.jp/rapdownload/rap2/gff_RAP2.tar.gz). The random insertion process was repeated 1,000 times. Replicates were used to estimate the expected number of insertions (and standard deviations) in each category under a random insertion model. The observed data of *mPing* insertions were compared to each replicate by a χ^2 test for the deviation from a random distribution.

RNA extraction and synthesis of Cy3- and Cy5-labelled cRNA. Total RNA was extracted from 7-day-old shoots of rice plants using an RNeasy Plant Mini Kit (Qiagen). Cy3- and Cy5-labelled cRNA was prepared from 400 ng of total RNA with Low RNA Input Linear Amplification Kit (Agilent Technologies) and Cy3- and Cy5-CTP (Perkin Elmer). Labelled cRNA was purified with RNeasy mini spin columns (Qiagen).

Microarray conditions and data analysis. A 44K Rice Oligo Microarray kit (Agilent Technologies) was used for microarray analysis. One microgram of Cy3-labelled cRNA was mixed with the same amount of Cy5-labelled cRNA and used for subsequent hybridization. Hybridization was carried out for 17 h with rotation at 60 °C. After washing, slides were scanned using a GenePix 4000A scanner (Axon Instruments) with 550 V and 680 V of PMT voltage for Cy3 and Cy5 detection, respectively, and quantified by Microarray Suite 2.0 (IPLab Spectrum Software, Scanalytics). Subsequent analysis was performed using GeneSpring 7 software (Agilent Technologies).

Genes that were up- or downregulated were selected as differentially transcribed between EG4 and NB. Signal intensity, amplitude of expression fluctuation and standard error of the mean F (F = the ratio of normalized data between NB and EG4) were also considered. Genes meeting the criteria were selected as follows: the average signal intensity of the control RNA in the experiment was within the range 53,103 to 13,107; the F of duplicate samples were all significantly higher or lower than 1 ($P < 0.01$).

Quantitative real-time PCR. Total RNA was prepared (for both rice and *Arabidopsis* seedlings) using an RNeasy Plant Mini Kit (Qiagen). Accumulation levels of the target transcripts were analysed by real-time PCR with a LightCycler 480 real-time PCR system (Roche Applied Science) by monitoring amplification with iQ SYBR Green Supermix (Bio-Rad) as described in the manufacturer's protocol.

Transformation of *Arabidopsis*. The *Ping* ORF1 promoter was cloned into the pDONR vector (Invitrogen) after PCR amplification. After sequencing, the promoter was transferred to pMDC 162 (ref. 30) by recombination, transformed into *Agrobacterium tumefaciens* (GV3103), and transformed into *Arabidopsis thaliana* ecotype Columbia by the floral dip method²⁹. Seedlings were germinated on Murashige and Skoog solid medium (0.2% phytogel, 1% sucrose), containing 50 $\mu\text{g ml}^{-1}$ kanamycin and 50 $\mu\text{g ml}^{-1}$ hygromycin. Staining with the β -Glucuronidase Reporter Gene Staining Kit (Sigma-Aldrich) was performed before and after cold treatment (4 °C for 2 h).

A transposon-induced epigenetic change leads to sex determination in melon

Antoine Martin¹, Christelle Troadec¹, Adnane Boualem¹, Mazen Rajab¹, Ronan Fernandez¹, Halima Morin², Michel Pitrat³, Catherine Dogimont³ & Abdelhafid Bendahmane¹

Sex determination in plants leads to the development of unisexual flowers from an originally bisexual floral meristem^{1,2}. This mechanism results in the enhancement of outcrossing and promotes genetic variability, the consequences of which are advantageous to the evolution of a species³. In melon, sexual forms are controlled by identity of the alleles at the andromonoecious (*a*) and gynoecious (*g*) loci⁴. We previously showed that the *a* gene encodes an ethylene biosynthesis enzyme, *CmACS-7*, that represses stamen development in female flowers⁵. Here we show that the transition from male to female flowers in gynoecious lines results from epigenetic changes in the promoter of a transcription factor, *CmWIP1*. This natural and heritable epigenetic change resulted from the insertion of a transposon, which is required for initiation and maintenance of the spreading of DNA methylation to the *CmWIP1* promoter. Expression of *CmWIP1* leads to carpel abortion, resulting in the development of unisexual male flowers. Moreover, we show that *CmWIP1* indirectly represses the expression of the andromonoecious gene, *CmACS-7*, to allow stamen development. Together our data indicate a model in which *CmACS-7* and *CmWIP1* interact to control the development of male, female and hermaphrodite flowers in melon.

In flowering plants, sex determination is a process that leads to the physical separation of male (stamens) and female (carpels) reproductive organs on different flowers (monoecy) or individuals (dioecy)^{1,2}. The development of unisexual flowers seems to have occurred several times independently, suggesting that plants may have evolved different sex-determination mechanisms⁶. In melon, monoecious (*AAGG*) and andromonoecious (*aaGG*) individuals bear male flowers on the main stem, and female or hermaphrodite flowers on axillary branches, respectively. Gynoecious (*AAGg*) and hermaphrodite (*aaagg*) individuals only bear female and hermaphrodite flowers, respectively⁴ (Supplementary Fig. 1). Cloning of the *a* gene showed that andromonoecy results from a mutation in the active site of the 1-aminocyclopropane-1-carboxylic acid synthase encoded by *CmACS-7*. Expression of the active enzyme inhibits the development of stamens and is not required for carpel development⁵. Unisexual females (gynoecy) arise in melon by the action of a recessive *g* allele, which leads to the development of female flowers instead of male flowers in the main stem of monoecious lines (Supplementary Fig. 1). To investigate the molecular basis leading to the transition from monoecy to gynoecy, and to understand how the combination of alleles at the *a* and *g* loci modify the sex of the plant, we undertook the positional cloning and the characterization of the gynoecious gene in the cross between the monoecious line PI 124112, and the gynoecious line Gynadou.

We cloned the *g* locus by first constructing a high-resolution genetic map, and used chromosome walking to develop a bacterial artificial chromosome (BAC) contig anchored to the genetic map (Fig. 1a). The *g*

locus was delimited to a single BAC clone, Bm102 from PI 124112, and sequence analysis of this clone revealed eight candidate genes within a region of 65 kilobase (kb) pairs (Fig. 1b). Sequence analysis of the Bm102 orthologous region in Gynadou revealed single-nucleotide polymorphisms (SNPs) that we typed in a segregating population of 12,660 individuals, further localizing the *g* locus to a 1.4-kb non-coding region. In Gynadou, we found an insertion of a DNA transposon of the *hAT* family⁷, henceforth termed *Gyno-hAT*, inside the mapped 1.4-kb region (Fig. 1b, Supplementary Fig. 2 and Supplementary Discussion). We screened a germplasm collection of 497 accessions of *Cucumis melo* from different regions of the world⁸ using a PCR marker specific to the insertion to test whether this transposon is completely associated to the *g* allele. All tested monoecious (*n* = 146) and andromonoecious (*n* = 347) accessions homozygous for the dominant *G* allele did not have this transposon insertion, whereas the hermaphrodite and gynoecious individuals containing the recessive *g* allele (*n* = 4) possessed the

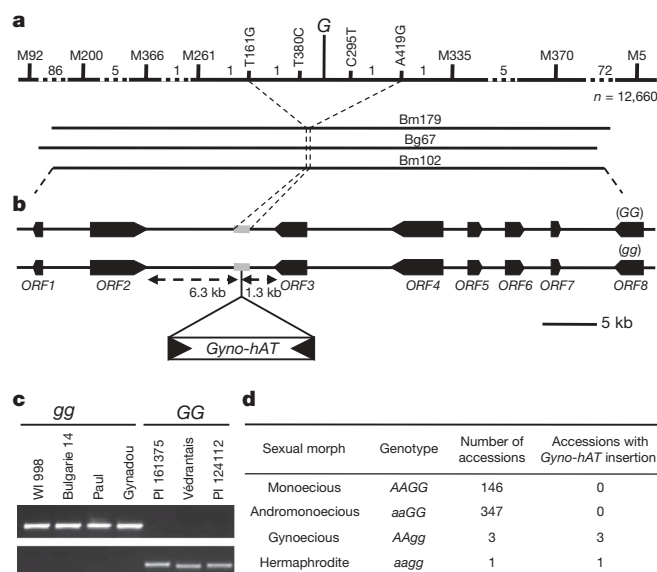


Figure 1 | Sex determination is due to the insertion of a DNA transposon at the gynoecious locus. **a**, Genetic and physical maps of the *g* locus. Numbers above the genetic map indicate the recombination events identified in the genotyping of 12,660 segregants. **b**, Representation of the 8 ORFs (broad arrows) found on BACs Bm102 and Bg67. The grey line indicates the 1.4-kb *G* locus minimal interval in PI 124112. In Gynadou (*gg*), this interval contains the insertion of *Gyno-hAT* transposon. **c**, PCR-based screen of presence (top) or absence (bottom) of the *Gyno-hAT* insertion in melon accessions homozygous for *g* or *G* alleles. **d**, PCR screening for *Gyno-hAT* insertion in 497 *C. melo* accessions, as described in **c**.

¹INRA-CNRS, UMR1165, Unité de Recherche en Génétique Végétale, 2 rue Gaston Crémieux, F-91057 Evry, France. ²Plateforme de Cytologie et d'Imagerie Végétale, Institut Jean Pierre Bourgin, INRA, 78026 Versailles Cedex, France. ³INRA, UR 1052, Unité de Génétique et d'Amélioration des Fruits et Légumes, BP 94, F-84143 Montfavet, France.

insertion, suggesting that insertion of *Gyno-hAT* is probably the cause of gynoecy in this species (Fig. 1c, d).

Transposable elements constitute a large fraction of plant genomes and are subjected to epigenetic modification^{8,9}, affecting the expression of linked genes, notably by the genomic spreading of DNA methylation^{10–13}. We examined the DNA methylation status of *Gyno-hAT* using the methylation-sensitive enzyme McrBC, and observed that it was heavily methylated (Fig. 2a). We searched for potential spreading of DNA methylation from *Gyno-hAT* to the three neighbouring genes, *ORF2*, *ORF3* and *ORF4* and found that only *ORF3*, located less than 1.5 kb from *Gyno-hAT*, was methylated (Fig. 2a). We compared the DNA methylation status of *ORF3* in plants containing a recombination event between *ORF3* and *Gyno-hAT* and observed that the *ORF3* allele from Gynadou was demethylated in the monoecious recombinant plant that lost *Gyno-hAT* (Supplementary Fig. 3). From this, we conclude that *Gyno-hAT* is required for induction and maintenance of *ORF3* DNA methylation. Consistent with epigenetic regulation, during somatic development gynoecious plants occasionally produce reverted flowers with stamens and reduced ovaries (Fig. 2b). We looked at the DNA methylation status of *ORF3* in those partially reverted flowers, and found that this phenotype was correlated with demethylation of the *ORF3* DNA sequence (Fig. 2c).

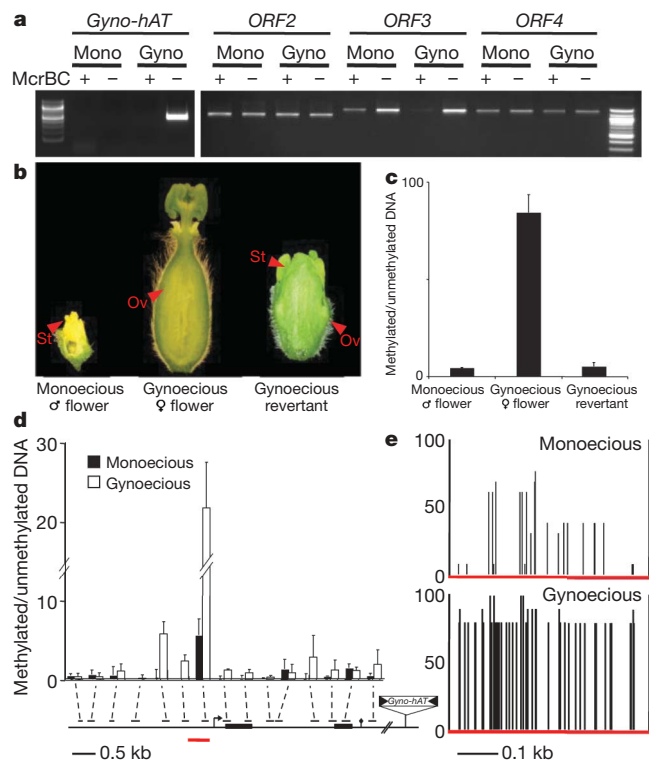


Figure 2 | The *Gyno-hAT* insertion leads to specific spreading of DNA methylation. **a**, McrBC-sensitive PCR analysis of *Gyno-hAT* transposon and the three closest ORFs in PI 124112 (Mono) and Gynadou (Gyno). + and – indicate whether genomic DNA was treated with McrBC before PCR amplification. The absence of PCR product in McrBC-treated samples indicates methylated DNA. **b**, Phenotypes of monoecious male flower, gynoecious female flower and gynoecious partially reverted flower. Ov, ovary; st, stamen. **c**, Quantitative McrBC-sensitive PCR on *ORF3* for the three flower types shown in **b**. **d**, Quantification of methylation along the *ORF3* genomic region, using McrBC-sensitive PCR. Correspondences between histogram bars and the analysed amplicons are indicated by dashed lines. The black arrow represents the transcription start site, black boxes represent the two *ORF3* exons. For **c** and **d**, data are means and s.d. ($n = 3$). **e**, Cytosine methylation analysis by bisulphite sequencing of the hypermethylated promoter region, located at –347 to –770 relative to *CmWIP1* ATG (underlined in red in **d**). The percentage of methylated cytosine is indicated by vertical bars.

To investigate this DNA-methylation pattern further, we analysed the whole *ORF3* genomic sequence using quantitative McrBC-PCR (Fig. 2d). The *ORF3* promoter region, upstream of the transcription start site, was strongly methylated in the gynoecious line (Fig. 2d). Weak DNA methylation was also observed in the *ORF3* transcribed region. Bisulphite sequencing confirmed hypermethylation of the *ORF3* promoter in the gynoecious line, attaining 97% of methylated cytosine in a CGN context (Fig. 2e and Supplementary Table 1). Mechanisms of DNA methylation spreading remain largely unknown, however it has been proposed that partially methylated DNA is a better substrate for *de novo* methylation^{10,14,15}. Consistent with this, bisulphite sequencing revealed low but significant DNA methylation in the *ORF3* promoter of the monoecious line, especially in a CGN context (Fig. 2e and Supplementary Table 1), which may explain the intensity as well as the polarization of *Gyno-hAT*-dependent *ORF3* methylation. Hypermethylation of the *ORF3* promoter was also confirmed in different genetic backgrounds recessive for the *g* allele, reinforcing the correlation between the presence of *Gyno-hAT*, increased methylation of *ORF3*, and gynoecy appearance in melon (Supplementary Fig. 4).

ORF3 encodes a C2H2 zinc-finger transcription factor of the WIP protein subfamily¹⁶. In *Arabidopsis thaliana*, two members of this plant-specific protein family, *TT1* and *NTT*, have been described, and shown to be involved in the development of female structures in the carpel^{16,17}. Annotation of WIP proteins showed a highly conserved C2H2 zinc-finger domain at the carboxy terminus¹⁶ and two conserved signatures at the very divergent amino terminus (Supplementary Fig. 5, Supplementary Table 2 and Supplementary Discussion). On the basis of this homology, we renamed *ORF3* *C. melo* *WIP1* (*CmWIP1*).

Because there is no efficient melon transformation protocol, the TILLING (targeting induced local lesions in genomes) approach was undertaken to confirm the role of *CmWIP1* in sex determination. We screened a monoecious genotype (AAGG) population of ethyl methane-sulphonate (EMS) mutants, and identified six mutations in the full *CmWIP1* coding region. One mutation was silent and five led to the missense mutations Ser25Phe, Leu77Phe, Pro193Leu, Gly242Arg and Ser306Phe (Supplementary Fig. 5). The changes Leu77Phe, Pro193Leu, Gly242Arg and Ser306Phe occur in highly conserved amino acid positions and may affect the function of the protein, whereas the Ser25Phe modification occurs at a non-conserved amino acid position (Supplementary Fig. 5). Backcrosses to the wild type showed, for more than 100 F_2 plants for each cross, that the Ser25Phe mutation, as well as the silent mutation, had no sexual phenotype. In contrast, plants homozygous for the Ser306Phe, Gly242Arg or Phe193Leu mutation showed a complete feminization of flowers in the whole plant, resulting in the formation of gynoecious individuals (Fig. 3a, b). The Leu77Phe mutant allele showed a weaker and intermediate phenotype, in which flowers from the main stem developed female traits such as ovaries and stigmas, as well as stamens (Fig. 3a, b). On the basis of the conversion of monoecy to gynoecy in different mutant plants, we concluded that *CmWIP1* is the gynoecious gene.

In unisexual flowers, the inappropriate sexual organs stop developing at stage 6, just after the elaboration of carpel and stamen primordia¹⁸. Quantitative real-time PCR analysis demonstrates that *CmWIP1* is highly expressed in male flowers from the monoecious line up to stage 6, and its expression subsequently drops off rapidly (Fig. 4a). *CmWIP1* expression was also weak in buds determined to develop a carpel (Fig. 4b). The spatial expression pattern of *CmWIP1* was examined using *in situ* hybridization. We observed that *CmWIP1* expression was mainly expressed in the carpel primordia of male flowers at stage 6 (Fig. 4c). No expression was detected in carpel primordia of buds determined to develop a carpel (Fig. 4d–f). As predicted, this low expression of *CmWIP1* in the gynoecious and hermaphrodite lines correlates with the hypermethylation of the *CmWIP1* promoter, and the expression in monoecious male flowers correlates with the weak methylation of the promoter. As *CmWIP1* expression leads to male flower development and the non-expression or the TILLING loss-of-function mutants

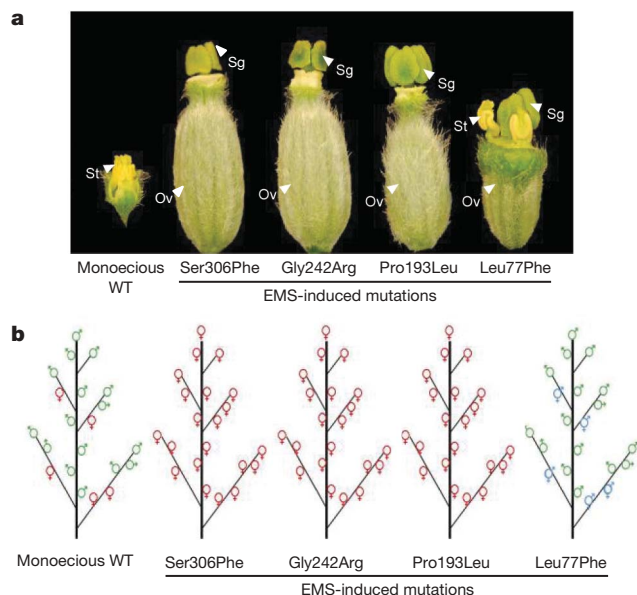


Figure 3 | *CmWIP1* is the gynoecious gene responsible for sex determination. **a**, Phenotypes of mature flowers in monoecious wild-type (WT) and EMS-induced mutants. Sepals and petals were removed to show the sexual organs. Sg, stigma. **b**, Schematic representation of flower sexual types observed in the whole plant of monoecious wild-type and EMS-induced mutants.

permit carpel development, we conclude that *CmWIP1* expression in carpel primordia prevents the development of female organs in the flower.

To integrate the function of the sex genes *CmWIP1* and *CmACS-7* in a single sex-determination model, we first compared the expression of both genes in wild-type plants. The expression of both genes is limited to carpel primordia, but with antagonistic patterns (Fig. 4b–h). We also analysed the expression of *CmACS-7* in *CmWIP1* TILLING loss-of-function mutants, and showed that *CmACS-7* is expressed when *CmWIP1* function is abolished (Fig. 4h–j). Together, these data indicate that *CmWIP1* acts upstream and is epistatic to *CmACS-7* in the sex-determination pathway. To investigate whether *CmWIP1* is a direct repressor of *CmACS-7* expression, we tested the physical interaction between *CmWIP1* protein and *CmACS-7* promoter, using a yeast one-hybrid system. The absence of interaction in yeast indicates that *CmWIP1* may indirectly repress the expression of *CmACS-7* (Supplementary Fig. 6). Together, these data indicate a model (Fig. 4k) in which the expression of *CmWIP1* causes the arrest of carpel development, but also indirectly, the repression of *CmACS-7* expression, allowing stamens to develop and leading to the formation of male flowers. The inactivation of *CmWIP1* by promoter hypermethylation in the g allele, or by induced mutation in the TILLING mutants, permits carpel development and also the expression of the stamens repressor *CmACS-7*, leading to the development of female flowers. In hermaphrodite flowers, stamens are also able to develop owing to the expression of a non-functional *CmACS-7* protein⁵. This integrated model leads to a complete understanding of the sex-determination processes in *C. melo*, the model plant for Cucurbitaceae. With regard

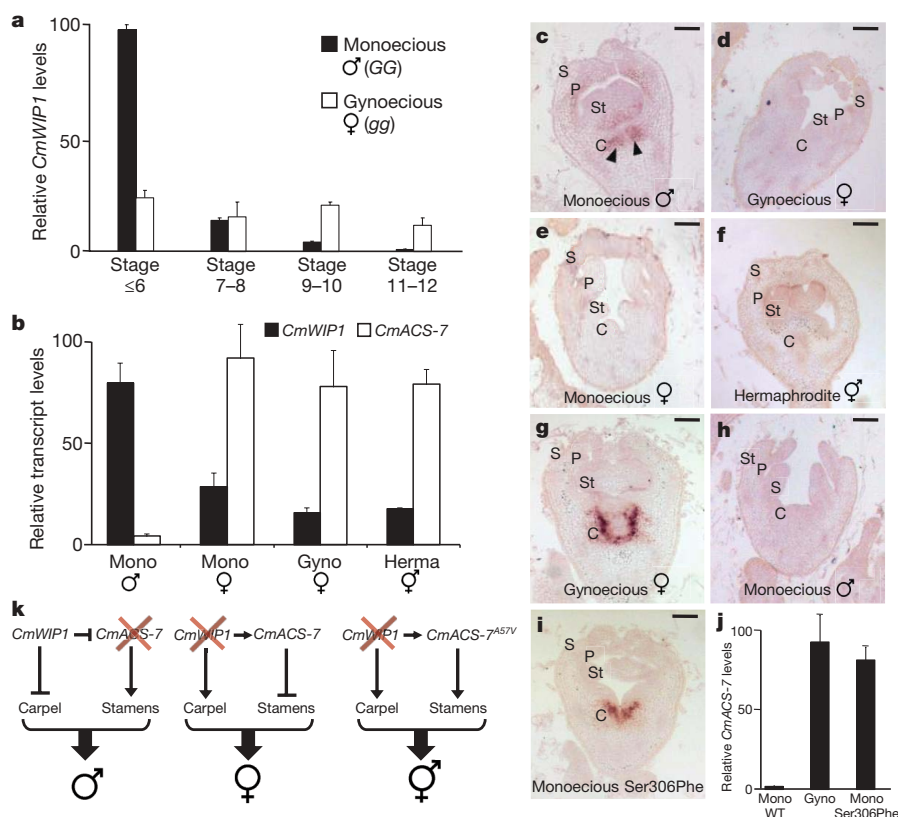


Figure 4 | Integration of *CmWIP1* and *CmACS-7* in a sex determination model. **a**, Quantitative real-time PCR (qPCR) of *CmWIP1* at different stages of flower development. **b**, qPCR of *CmWIP1* and *CmACS-7* in flower buds, collected up to stage 6, from monoecious male or female flowers, gynoecious female flowers and hermaphrodite flowers. **c–f**, *In situ* localization of *CmWIP1* mRNA in monoecious (male, **c** or female, **e**), gynoecious (d) and hermaphrodite (f) flowers. *CmWIP1* transcripts were detected in carpel primordia of male flowers at stage 6 (black arrows, **c**). No transcripts were

detected in any female or hermaphrodite flowers. **g–i**, *In situ* localization of *CmACS-7* in flower buds from the gynoecious line (**g**), TILLING Ser306Phe mutant plant (**i**) and the wild-type control plant (**h**). **j**, qPCR of *CmACS-7* in flower buds from monoecious wild-type, gynoecious and TILLING Ser306Phe mutant plants. For qPCR, the mean \pm s.d. of three biological replicates is shown. Scale bars, 200 μ m. C, carpels; P, petals; S, sepals. **k**, Model of the sex-determination processes in melon integrating *CmWIP1* and *CmACS-7* genetic and functional information.

to the nature of the mutation at the gynoeious locus, although recent reports show that epimutations can affect plant growth and development^{19–21}, this naturally occurring transposon-induced epigenetic change may be especially important for environmental adaptation and evolution, as it affects an essential trait for plant reproduction. Recently, it emerges that the higher plasticity of an epiallele can be expected to provide substantial benefit for the adaptation to stress and variations of environmental conditions²⁰. Therefore, the consequence of this epigenetic alteration may have significance for the long-term evolution of this species.

METHODS SUMMARY

Genetic dissection of the *g* locus. We crossed the monoecious cultivar PI 124112 (AAGG) and the gynoeious cultivar Gynadou (AAGg), and backcrossed the F₁ plants with Gynadou. To identify markers linked to the *g* locus, we screened 1,024 MseI/PstI amplified fragment length polymorphism (AFLP) primer combinations using bulked DNA samples of 12 gynoeious and 12 monoecious plants²². Thirteen AFLP markers were identified and mapped relative to the *g* locus in a segregating population of 300 individuals. To identify recombination events close to the *g* locus, we genotyped DNA samples from 12,660 plants in the mapping population with the *g*-flanking markers M92 and M5. One-hundred-and-sixty-eight plants carrying recombination events in the M92–M5 region were identified and phenotyped for the sexual morph.

McrBC-based methylation assay. McrBC-based methylation analysis was performed as described previously²³. Genomic DNA was digested overnight with McrBC in parallel with a mock digestion, and the equivalent of 15 ng was used for PCR. Samples were normalized with *UBIQUITIN10* unmethylated gene. Unmethylated versus methylated DNA was calculated with the formula $2^{\Delta C_t}$ where C_t is the threshold cycle; ΔC_t is the difference of C_t between digested and mock samples. An unmethylated sample gives a ΔC_t value close to 0, and a ratio unmethylated versus methylated value close to 1. Semi-quantitative PCR consisted of 28 PCR cycles.

Reverse-genetic and TILLING experiments. Allelic *CmWIP1* mutants were isolated from a pool of EMS-induced mutant seeds obtained from a monoecious population (GG) of 3,000 M2 families. Genomic DNA isolation, PCR amplification and detection of mutations were performed as previously described²⁴. Homozygous mutants were backcrossed to the wild-type, and the segregation of the mutations with the phenotypes was followed in more than 100 F₂ plants.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 June; accepted 9 September 2009.

1. Tanurdzic, M. & Banks, J. A. Sex-determining mechanisms in land plants. *Plant Cell* **16** (suppl), S61–S71 (2004).
2. Dellaporta, S. L. & Calderon-Urrea, A. Sex determination in flowering plants. *Plant Cell* **5**, 1241–1251 (1993).
3. Dellaporta, S. L. & Calderon-Urrea, A. The sex determination process in maize. *Science* **266**, 1501–1505 (1994).
4. Poole, C. F. & Grimball, P. C. Inheritance of new sex forms in *Cucumis melo* L. *J. Hered.* **30**, 21–25 (1939).
5. Boualem, A. et al. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* **321**, 836–838 (2008).
6. Ainsworth, C., Parker, J. & Buchanan-Wollaston, V. Sex determination in plants. *Curr. Top. Dev. Biol.* **38**, 167–223 (1998).
7. Kempken, F. & Windhofer, F. The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma* **110**, 1–9 (2001).

8. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nature Rev. Genet.* **8**, 272–285 (2007).
9. Weil, C. & Martienssen, R. Epigenetic interactions between transposons and genes: lessons from plants. *Curr. Opin. Genet. Dev.* **18**, 188–192 (2008).
10. Arnaud, P., Goubely, C., Pelissier, T. & Deragon, J. M. SINE retrotransposons can be used *in vivo* as nucleation centers for *de novo* methylation. *Mol. Cell. Biol.* **20**, 3434–3441 (2000).
11. Sun, F. L. et al. *cis*-Acting determinants of heterochromatin formation on *Drosophila melanogaster* chromosome four. *Mol. Cell. Biol.* **24**, 8210–8220 (2004).
12. Martienssen, R. A., Doerge, R. W. & Colot, V. Epigenomic mapping in *Arabidopsis* using tiling microarrays. *Chromosome Res.* **13**, 299–308 (2005).
13. Saze, H. & Kakutani, T. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J.* **26**, 3641–3652 (2007).
14. Christman, J. K., Sheikhejad, G., Marasco, C. J. & Sufrin, J. R. 5-Methyl-2'-deoxycytidine in single-stranded DNA can act in *cis* to signal *de novo* DNA methylation. *Proc. Natl Acad. Sci. USA* **92**, 7347–7351 (1995).
15. Jones, P. A. & Takai, D. The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068–1070 (2001).
16. Sagasser, M., Lu, G. H., Hahlbrock, K. & Weisshaar, B. A. *thaliana* TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. *Genes Dev.* **16**, 138–149 (2002).
17. Crawford, B. C., Ditta, G. & Yanofsky, M. F. The *NTT* gene is required for transmitting-tract development in carpels of *Arabidopsis thaliana*. *Curr. Biol.* **17**, 1101–1108 (2007).
18. Bai, S. L. et al. Developmental analyses reveal early arrests of the spore-bearing parts of reproductive organs in unisexual flowers of cucumber (*Cucumis sativus* L.). *Planta* **220**, 230–240 (2004).
19. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–161 (1999).
20. Kalisz, S. & Purugganan, M. D. Epialleles via DNA methylation: consequences for plant evolution. *Trends Ecol. Evol.* **19**, 309–314 (2004).
21. Manning, K. et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature Genet.* **38**, 948–952 (2006).
22. Bendahmane, A., Kanyuka, K. & Baulcombe, D. C. The *Rx* gene from potato controls separate virus resistance and cell death responses. *Plant Cell* **11**, 781–792 (1999).
23. Pontier, D. et al. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
24. Triques, K. et al. Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. *Plant J.* **51**, 1116–1125 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Bradley, M. Purugganan, P. Laufs, J. C. Palauqui and colleagues from Unité de Recherche en Génomique Végétale (URGV) for discussions and comments on the manuscript, and F. Teixeira and V. Colot for helpful assistance on DNA methylation assays. This work was supported by GAP department in INRA and by a grant from Genopole (A.M.).

Author Contributions A.Be., C.D. and M.P. initiated the project. A.M., C.D. and A.Be. designed the experiments. A.M., C.T., A.Bo., M.R., R.F. and H.M. performed the experiments, for which the data were analysed by A.M., C.D. and A.Be. A.M. and A.Be. wrote the paper, which was commented on and improved by all the authors.

Author Information Sequences have been deposited in GenBank under accession codes GQ870274 and GQ870275. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.Be. (bendahm@evry.inra.fr).

METHODS

Plant material. A population between the monoecious cultivar PI 124112 (AAGG) and the gynoeious cultivar Gynadou (AAGg) was developed, and backcrossed the F₁ plants with Gynadou. One month after sowing, plants were transferred in soil, under field-like conditions to allow optimal root growth. Plants were grown in a greenhouse in Montfavet (France) during the summer (from the beginning of July to the end of August), under natural light conditions. Plants were genotyped around the two-leaf stage. Sex was phenotyped in the two-months after transferring the plants to the greenhouse. Phenotyping were performed on the main axis and on branches. Revertant lines were grown in a growth chamber under conditions of long days (16 h light at 26 °C, 8 h dark at 19 °C). Plants were watered with a solution of 3.5:3.1:8.6 N:P:K.

Construction and screening of BAC libraries. To construct a physical map, overlapping BAC clones spanning the *g* locus were identified and anchored to the genetic map. In this analysis, the *g* locus was physically delimited to a single BAC clone, Bm102, derived from PI 124112 BAC library⁵. To identify Bm102 orthologous region in Gynadou, we constructed a second melon BAC library from Gynadou as described previously⁵. The library consists of 50,000 clones and represents the haploid melon genome at least 10 times over. Using a systematic PCR-based procedure, the library was screened with *g* flanking markers M261 and M335, and the BAC clone Bg67 was identified.

Sequencing and annotation of the melon Bm102 and Bg67. Bm102 and Bg67 were sequenced using a shotgun sequencing procedure. Sequence contigs were assembled with UNIX versions of the Staden program package²⁵. After assembly, two forms of gene annotation were performed: (1) *ab initio* gene prediction with GENSCAN (<http://genes.mit.edu/GENSCAN.html>), GeneMark.hmm (<http://exon.gatech.edu/GeneMark/eukhmm.cgi>) and FGENESH (<http://www.softberry.com/berry.phtml>); and (2) homology to known genes with BLAST on GenBank (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) and Genseq (<http://www.plantgdb.org/cgi-bin/GeneSequer/PlantGDBgs.cgi>).

Sequence analysis of the M261–M335 genomic interval. Sequence analysis of Bm102 and Bg67 inserts showed SNPs that we mapped relative to the recombination events identified in a segregating population of 12,660 individuals, further reducing the *g* locus to a 1.4-kb non-coding region. In Gynadou, an insertion of 8 kb, corresponding to a DNA transposon of the *hAT* family, that we named *Gyno-hAT*, was found inside the 1.4-kb region.

Bisulphite sequencing. Bisulphite sequencing was performed using the MeathylEasy kit (Human Genetic Signatures). Unmethylated cytosine in genomic DNA (500 ng) were converted to uracil and amplified by PCR as the manufacturer recommends. Oligonucleotides used for bisulphite sequencing are listed in Supplementary Table 3. Around ten clones were sequenced to estimate the frequency of methylation for each cytosine in the amplified DNA fragment. Sequences were then analysed using Cymate software²⁶.

RACE experiments. Total RNA was extracted from apices using Trizol reagent (Invitrogen). Amplifications of full-length cDNA 5' and 3' ends were performed using the RNA ligase-mediated method of GeneRacer kit (Invitrogen). PCR and nested PCR were performed according to the manufacturer advices. Oligonucleotides used for RACE PCR are listed in Supplementary Table 3. Amplified bands were gel-purified, cloned and sequenced.

Quantitative PCR analysis. Total RNA was extracted from flowers using Trizol reagent (Invitrogen), and 1 µg was reverse-transcribed using the Superscript III enzyme (Invitrogen). Transcripts levels were quantified using the ABI PRISM 7900 SDS quantitative PCR (qPCR) system (Applied Biosystems). The qPCR reactions contained 5 µl qPCR MasterMix Plus for SYBR Green I w/o UNG (Eurogentec), the equivalent of 10 ng total RNA and 200 nM of each primer, and were run in triplicates, using the following two-step cycling program: 95 °C for 15 s, 60 °C for 1 min for 40 cycles. qPCR specificity was checked at each run with the dissociation curve, at temperature ranging from 94 °C to 60 °C. Data from qPCR were analysed using the $2^{-\Delta\Delta C_t}$ method²⁷, where C_t is the threshold cycle. The housekeeping gene *ACTIN* was used to normalize the qPCR output. Oligonucleotides used for qPCR are listed in Supplementary Table 3.

In situ hybridization analysis. *In situ* hybridizations were performed as described previously⁵. The complete coding sequence of *CmWIP1* or *CmACS-7* was used to generate the antisense labelled probes. Flowers from stages 1–6, 7–8, 9–10 and 11–12 were fixed and used for hybridizations. *CmWIP1* expression was mainly detected in carpel primordial, and a signal was also observed in the adaxial side of stamens and petals (Fig. 4c).

Protein–promoter interaction in yeast. The potential physical interaction between *CmWIP1* protein and the *CmACS-7* promoter was investigated using the Matchmaker one-hybrid kit (Clontech Laboratories). A sequence 2 kb upstream of the *CmACS-7* start codon of was introduced into pHIS2.1 vector that carries the tryptophan synthesis gene *Trp1* as a transformation marker. The coding sequence of *CmWIP1* was introduced into pGADT7-rec2 vector that carries the leucine synthesis gene *Leu2* as a transformation marker. Interaction between a protein and a promoter sequence stimulates the expression of the *His3* gene, enabling the yeast to grow on minimal media lacking histidine. The two constructions were used to co-transform the Y187 yeast strain. The interaction was verified on selective medium in comparison to a positive control provided by the kit.

25. Staden, R., Beal, K. F. & Bonfield, J. K. The Staden package, 1998. *Methods Mol. Biol.* **132**, 115–130 (2000).
26. Hetzl, J., Foerster, A. M., Raidl, G. & Scheid, M. O. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**, 526–536 (2007).
27. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method. *Methods* **25**, 402–408 (2001).

Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization

Jason M. Crawford^{1*}, Tyler P. Korman^{4*}, Jason W. Labonte¹, Anna L. Vagstad¹, Eric A. Hill¹, Oliver Kamari-Bidkorphe⁵, Shiou-Chuan Tsai^{4,5,6} & Craig A. Townsend^{1,2,3}

Polyketides are a class of natural products with diverse structures and biological activities. The structural variability of aromatic products of fungal nonreducing, multidomain iterative polyketide synthases (NR-PKS group of IPKs) results from regiospecific cyclizations of reactive poly- β -keto intermediates^{1–3}. How poly- β -keto species are synthesized and stabilized, how their chain lengths are determined, and, in particular, how specific cyclization patterns are controlled have been largely inaccessible and functionally unknown until recently⁴. A product template (PT) domain is responsible for controlling specific aldol cyclization and aromatization of these mature polyketide precursors, but the mechanistic basis is unknown. Here we present the 1.8 Å crystal structure and mutational studies of a dissected PT monodomain from PksA, the NR-PKS that initiates the biosynthesis of the potent hepatocarcinogen aflatoxin B₁ in *Aspergillus parasiticus*. Despite having minimal sequence similarity to known enzymes, the structure displays a distinct ‘double hot dog’ (DHD) fold. Co-crystal structures with palmitate or a bicyclic substrate mimic illustrate that PT can bind both linear and bicyclic polyketides. Docking and mutagenesis studies reveal residues important for

substrate binding and catalysis, and identify a phosphopantetheine localization channel and a deep two-part interior binding pocket and reaction chamber. Sequence similarity and extensive conservation of active site residues in PT domains suggest that the mechanistic insights gleaned from these studies will prove general for this class of IPKs, and lay a foundation for defining the molecular rules controlling NR-PKS cyclization specificity.

Aflatoxin B₁ (Fig. 1a, 3) biosynthesis is initiated by the NR-PKS PksA, which accepts a hexanoyl starter unit from a dedicated fungal fatty acid synthase (FAS) and extends it through seven iterative, malonyl-derived ketide extensions to norsolorinic acid anthrone ((1), noranthrone) (Fig. 1a)⁵. Application of the Udvardy–Merski algorithm (UMA) afforded the unanticipated insight that PksA was composed of two unrecognized domains, in addition to the known β -ketoacyl synthase (KS), malonyl-CoA:ACP transacylase (MAT), acyl-carrier protein (ACP) and thioesterase (TE) domains⁶. Deconstruction experiments on PksA established that the two domains with previously unknown function were a starter unit:ACP transacylase (SAT)⁷ and a product template (PT)⁴. The PT domain was necessary for binding the fixed-length linear intermediate 6 and catalysing the

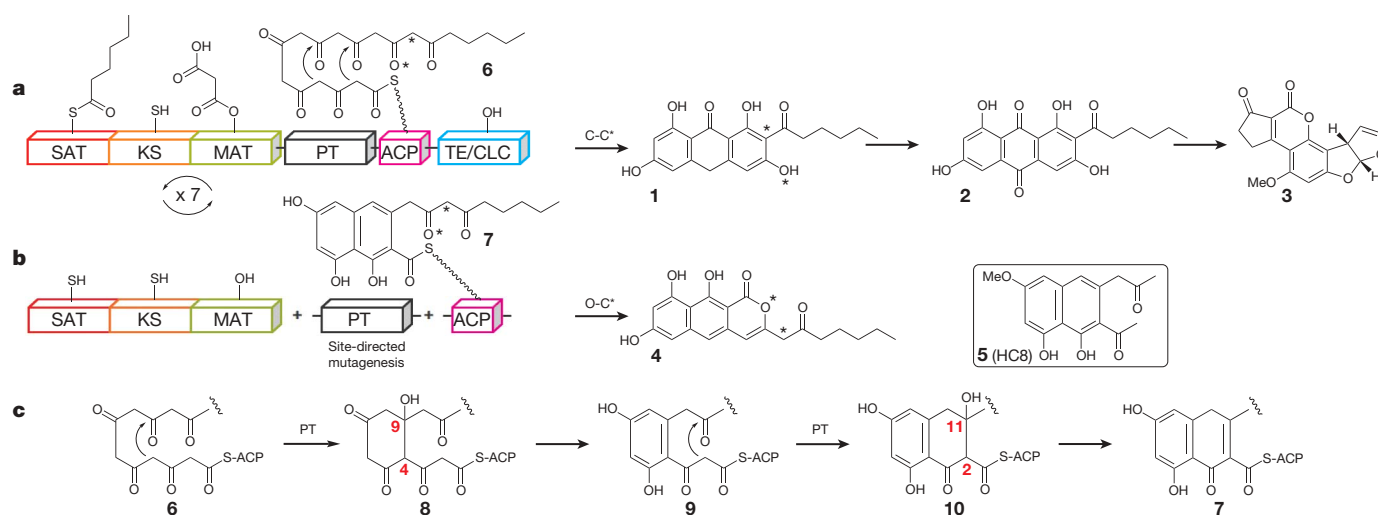


Figure 1 | Biosynthesis of norsolorinic acid anthrone (1) by PksA. **a**, Native catalytic cycle. The SAT domain in PksA selects a hexanoyl starter unit. The MAT domain loads the free ACP with malonyl units. After seven successive condensation events with malonyl-ACP catalysed in KS, the linear ACP-bound polyketide 6 is cyclized (C4–C9 and C2–C11 cyclization events) and aromatized in the product template (PT) domain to give the bicyclic intermediate 7. The TE catalyses C–C cyclization to release anthrone 1,

which undergoes oxidation to the anthraquinone norsolorinic acid (2), to initiate the complex biosynthetic pathway to aflatoxin B₁ (3). Asterisks denote the C and O nucleophiles involved in on-path and off-path cyclizations, respectively. **b**, In the absence of TE, the PT product 7 undergoes spontaneous O–C cyclization to norpyrone (4). The PT monodomain structure was solved with intermediate/product mimic 5 (HC8). **c**, Conversions from 6 to 7 via intermediates 8, 9 and 10.

¹Department of Chemistry, ²Department of Biology, ³Department of Biophysics, Johns Hopkins University, Maryland 21218, USA. ⁴Department of Molecular Biology and Biochemistry, ⁵Department of Chemistry, ⁶Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, USA.

*These authors contributed equally to this work.

efficient, stepwise aldol cyclizations (**8** and **10**) and their dehydrations (**7** and **9**) to yield the bicyclic ACP-thioester intermediate **7**, which undergoes final Claisen/Dieckmann cyclization in the TE domain to release noranthrone (**1**) (Fig. 1a, c)⁴. When the TE domain was absent, naphthopyrone (**4**) was produced by self-condensation of the stalled bicyclic PT product **7** (Fig. 1b).

We crystallized PT in $P2_12_1$ and $P4_12_1$ space groups (Supplementary Table 1) and solved the PT crystal structure by multi-wavelength anomalous dispersion (MAD) using selenomethionine-derivatized PT followed by phase extension with the native data to 1.8 Å. The crystal structures from both crystal forms showed distinct electron density for palmitate (from *Escherichia coli* purification), which could be removed using lipid-binding resin followed by immediate incubation with the product analogue HC8 (**5**, Fig. 1b) with PT. The HC8-bound PT structure was solved by molecular replacement to 1.9 Å using the palmitate-bound structure. The PT domain (~39 kDa) was isolated as a homodimer in solution and exists as a dimer in the crystal structure (Fig. 2a).

The PT primary sequence bears little similarity to any functionally characterized protein, except similarly placed domains in other fungal NR-PKSs. However, the PT crystal structure displays a DHD fold (Fig. 2a)⁸ that is a variant of those observed in the dehydrase (DH) domains of animal FASs and bacterial modular PKSs (Fig. 2b)^{9,10}. Two DHD monomers associate via a PT-specific sequence insertion (Supplementary Figs 1 and 2), leading to a molecule with dimensions $116 \times 53 \times 32$ Å. On the basis of sequence similarity, the DHD fold and dimer interface are conserved in the PT domains from fungal NR-PKSs.

The palmitate- and HC8-bound PT structures identify a deep pocket that extends 30 Å from the surface to the bottom (Fig. 3a) with well defined electron density for palmitate (PLM, Fig. 3b) or HC8 (**5**, Fig. 3c). The two crystal structures have a root mean squared deviation of 0.38 Å, indicating that little conformational change occurs in the rigid PT binding pocket in the presence of either linear or bicyclic compounds (except for Asn 1568, see later). The internal pocket can be visually divided into three regions (Fig. 3a): (1) the phosphopantetheine (PPT)-binding region extends 14 Å from the protein surface into the pocket. While the downstream ACP domain is transiently docked to the PT surface, this region is proposed to bind the ~18 Å PPT arm that delivers the substrate **6** into the pocket. (2) The cyclization chamber (8×13.5 Å) can accommodate two aromatic rings (Fig. 3a, c) and contains the proposed catalytic dyad (Asp 1543/His 1345) positioned appropriately to initiate regiospecific cyclizations. The PT HC8-bound structure positioned HC8 near this chamber (detailed in Supplementary Information). The central location of the cyclization chamber suggests that the substrate binds in an extended conformation with a pre-cyclization 'kink' in the middle of the polyketide (Figs 3a and 4). Notably, hydrophilic residues and a network of

crystallographic water molecules constitute a distinctly 'wet' side to the reaction chamber. In contrast, the orthogonal pocket proposed to accommodate the aldol cyclization products is largely hydrophobic and 'dry'. (3) The hydrophobic hexyl-binding region (6×6 Å) lies at the bottom and is perfectly adapted to accept the substrate hexanoyl starter unit. The tail of the 16-carbon palmitate is positioned directly in this region, exemplifying its tight aliphatic binding capabilities. In PTs that do not bind a hexyl group, Gly 1491 at the bottom of the hexyl-binding region becomes a bulky residue and closes off the pocket (Supplementary Fig. 1). The above observations strongly support the view that the hexyl- and PPT-binding regions anchor the ends of the extended ACP-linked poly- β -keto substrate, and the cyclization chamber accommodates two regiospecific cyclizations.

The PT active site has three features distinct from other DHDs: (1) the catalytic histidine (His 1345) uses its N δ instead of N ϵ as the basic nitrogen (Fig. 4 and Supplementary Fig. 3). (2) The catalytic Asp 1543 polarizes His 1345. In DHs and hydratases, the active site His is polarized by a neighbouring backbone carbonyl oxygen, not by the Asp^{11,12}. (3) The oxyanion hole exists in an altered location compared to known DHDs, where the second half of the oxyanion hole typically is the backbone N-H of a glycyl residue before the 'hot dog' helix^{11,12}. This glycine is replaced by a proline (Pro 1355) in PksA—the only residue incapable of providing an amide hydrogen bond. These differences suggest that PksA PT has evolved substantially from a simple DH or hydratase.

The palmitate- and HC8-bound structures, docking simulations (detailed in Supplementary Information) and comparisons to related enzymes support the following proposed mechanism for PT-catalysed first- and second-ring cyclization and aromatization (Fig. 4 and Supplementary Fig. 3): the ACP-bound linear substrate **6** binds to the PT substrate pocket in an extended conformation, as exemplified in the palmitate-bound structure and simulations. The lack of a salt bridge between the carboxylate of palmitate and the catalytic His implies that His 1345 is present as histidine (not histidinium) and functions as the key catalytic base. Acting together with Asp 1543 (hydrogen-bonded to, and oriented by, Gln 1547), His 1345 deprotonates C4, leading to the enolate intermediate stabilized by the backbone N-H of Asn 1568. Collapse of the enolate and subsequent C4 aldol addition to C9 carbonyl (C4–C9 cyclization event) leads to the first ring cyclization product **8**, with the network of water molecules bound to Ser 1356, Asp 1543, Asn 1568 and Thr 1546 stabilizing the oxyanion (altered PT oxyanion site). The extent of enolization in β -diketones is known to be highly dependent on the polarity of the surrounding medium¹³. 2,4-Pentanedione, for example, exists in its diketo form **5** or **6**:1 in water, but is almost entirely enolized in cyclohexane (97–98%)¹⁴. The role of the 'wet' side of the reaction chamber becomes evident with its water network ensuring that the C9 carbonyl

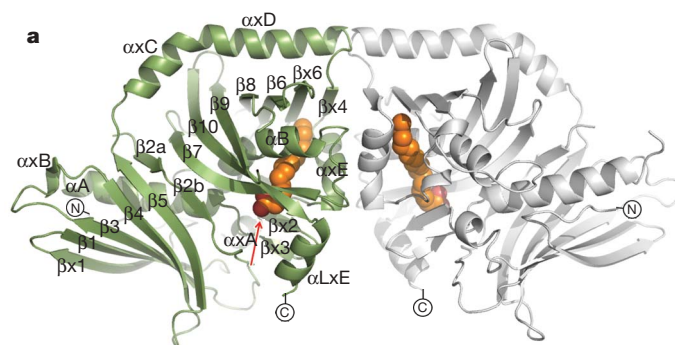
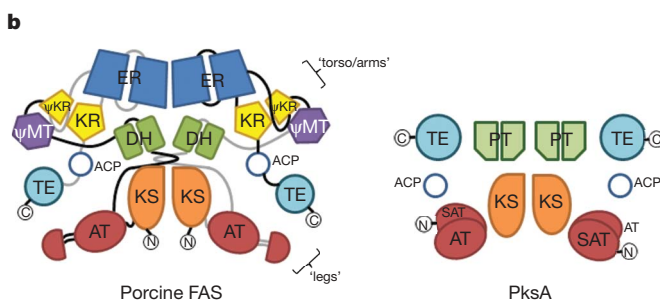


Figure 2 | The crystal structure of the PT domain from PksA. PT contains a modified double hot dog (DHD) fold. **a**, Ribbon representation of the PT dimer with the interface between monomer A (green, with secondary structural elements labelled) and the entrance to the pocket indicated with a red arrow) and monomer B (grey) stabilized by the elongated helices. Palmitate is shown in orange spheres. The α -helix and β -strand order is



indicated on the left monomer unit. **b**, Proposed domain organization of PksA (right) and comparison to the structure of animal FAS (left). ACP, acyl-carrier protein; AT, acyl transacylase; DH, dehydrase; ER, enoyl reductase; KR, ketoreductase; KS, ketoacyl synthase; Ψ MT, pseudo-methyltransferase; SAT, starter unit:ACP transacylase; TE, thioesterase.

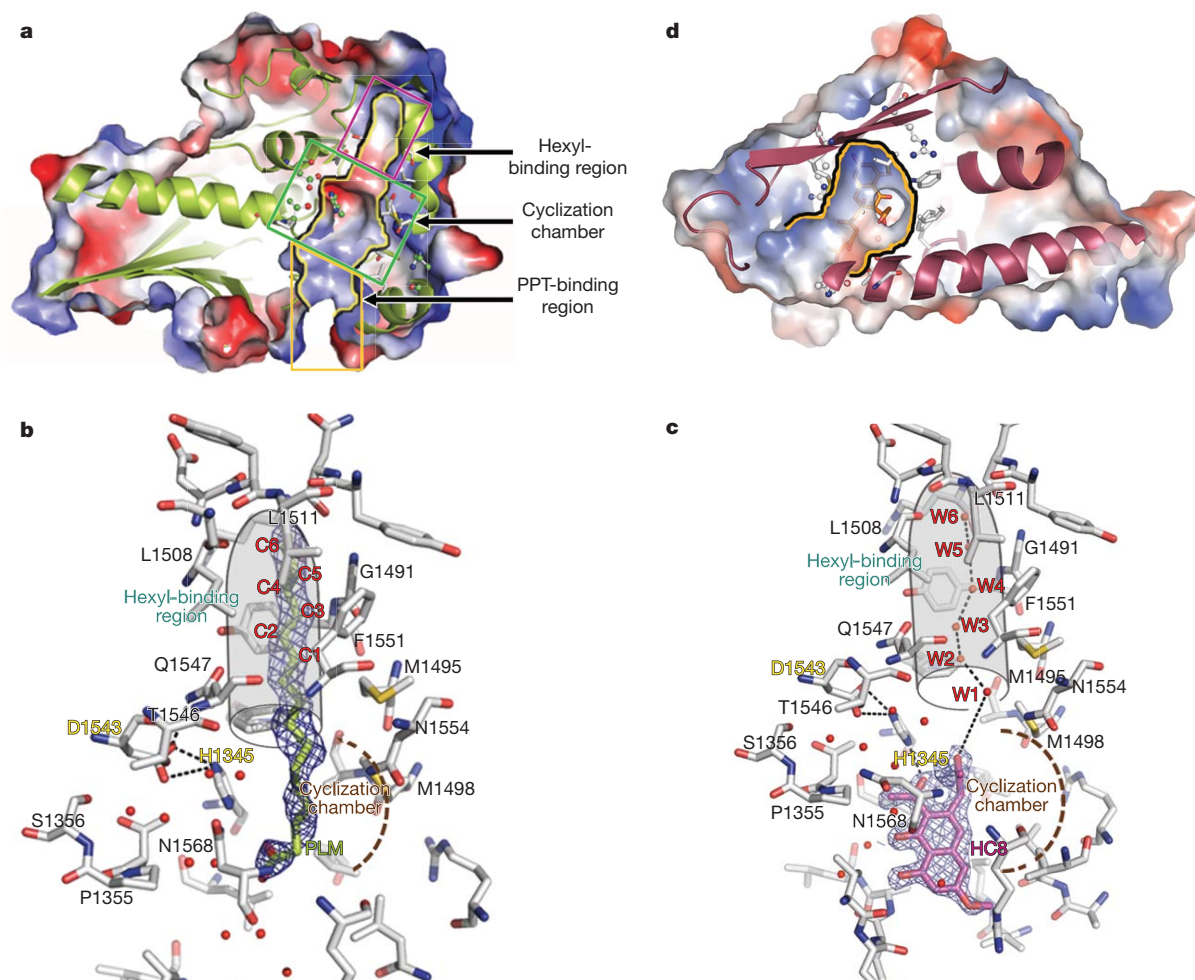


Figure 3 | PT structures with bound palmitate or substrate analogue HC8. **a**, Surface representation of the PT monomer with the pocket shown (yellow outline). The hexyl-binding region, cyclization chamber and PPT-binding region are highlighted. **b**, The location and orientation of the hexyl-binding region is shown as a grey tube. The cyclization chamber location is indicated with a brown dashed line. The simulated annealing (SA) omit map calculated after omitting the bound palmitate is contoured at 1.3σ as blue mesh around

the modelled palmitate ($2F_o - F_c$). **c**, Same as **b**, except with HC8 bound. The SA omit map contoured at 1.5σ was calculated after omitting the bound HC8 and is shown as blue mesh. **d**, The bacterial Tcm ARO/CYC crystal structure reveals a reaction chamber (orange outline) shallower and broader than the fungal PksA PT that accommodates its C_{20} polyketide substrate in a 'hairpin' conformation leading to C9–C14 first ring cyclization.

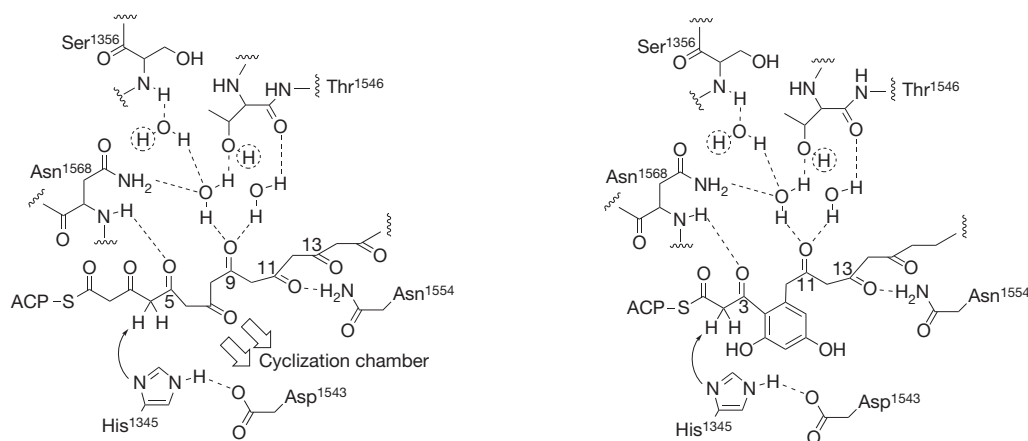


Figure 4 | Proposed mechanism of cyclizations catalysed by the PT domain. For clarity of presentation, liberties have been taken with the relative placement of active-site residues, but the principal interactions with the fully extended polyketide intermediate (left) are depicted, notably, to ordered crystallographic water molecules and a slight 'kink' towards the

cyclization chamber. Hydrogens indicated by dotted circles form hydrogen bonds to acceptors on the catalytic Asp 1543. The His 1345/Asp 1543 catalytic dyad acts as the catalytic base, leading to enolate formation at C4 to initiate first ring cyclization (C4–C9). The second ring is formed in the same manner as the first to achieve C2–C11 closure (Supplementary Fig. 3).

is both in its electrophilic keto tautomeric form and polarized through hydrogen bonding to activate it for intramolecular aldol cyclization, whereas the conserved Asn 1554 could anchor C11 or C13 by hydrogen bonding. First ring aromatization could occur with a loss of hydroxide (E1cb) similar to scytalone dehydratase¹⁵ and enoyl-CoA dehydratase¹⁶, but with a larger thermodynamic driving force. The first aromatized ring then translocates into the cyclization chamber, and the second ring (C2–C11) cyclization and aromatization occurs analogously to the first ring cyclization event. We take the view exemplified by recent experiments with *Mycobacterium tuberculosis* FabH that product release and substrate binding occur in an ‘open’ form of the protein, whereas catalysis takes place in the ‘closed’ form apparent in the crystal structures¹⁷.

The importance of the proposed active site residues was verified by independently mutating His 1345, Asp 1543, Gln 1547 and Asn 1554 to Ala, which resulted in no detectable activity (detailed in Supplementary Discussion and Supplementary Figs 3 and 4). The Thr 1546 backbone amide carbonyl and side-chain hydroxyl are each hydrogen bonded to one of the two ordered water molecules proposed to stabilize the substrate electrophilic centre of the aldol cyclization reactions (Fig. 4). Mutation of this residue to Ala disrupts one of these interactions and activity falls to approximately 40%. In the HC8-bound structure, the Asn 1568 side-chain amide is hydrogen bonded to one of these water molecules, whereas in the palmitate structure it moves to hydrogen bond to the fatty acid carboxylate—but the ordered water remains in place. Here mutation to Ala slightly raises production of the PT-released product **4** (approximately 120%), presumably by reducing steric constraints in the active site where substrate movement is occurring during the cyclization steps. Furthermore, Gly1491Leu represents a bulky mutation to block the hexyl-binding pocket, which also resulted in no apparent activity. Sequence alignments suggest that the catalytic mechanism for PT domains in fungal PKSs is conserved (Supplementary Fig. 1). As expected, the largest variations lie within the cyclization chamber in keeping with their proposed function.

To evaluate the role of PT in the global domain organization of type I nonreducing IPKSs, the oligomerization states of independent mono-, di- and tridomains were determined by native PAGE (detailed in Supplementary Table 2), and the results support a structural organization of PksA with similarity to animal FAS⁹ (Fig. 2b). The above analyses coupled with previous phylogenetic results¹⁸ suggest that the dimeric PT domain evolved from ancient DH domains in reducing PKSs and would be positioned at the DH locus in animal FAS.

Two distinct cyclization patterns for aromatic polyketides from fungi and *Streptomyces* have been identified². The S-pattern, in which the first ring is cyclized between C7–C12 or C9–C14, is observed in type II PKSs from *Streptomyces*. Alternatively, the F-pattern denotes first ring cyclization between C4–C9⁴, C2–C7¹⁹, or C6–C11², which is observed exclusively in fungal aromatic IPKSs (Supplementary Fig. 5). Recently, the molecular basis of the S-folded cyclization pattern (C9–C14) was elucidated by solving the tetracenomycin (Tcm) ARO/CYC crystal structure (Fig. 3d) and conducting pocket residue mutagenesis²⁰. In the present work, we have shown that the PT domain is capable of driving first- and second-ring cyclization from a pre-assembled, fixed-length poly- β -keto chain to promote the F-folded cyclization pattern. The Tcm ARO/CYC structure exhibits a helix-grip fold (Supplementary Fig. 2)²⁰ that has topology similar to the DHD fold of PT. The crystal structures, mutagenesis experiments and earlier mass spectrometric results⁴ support the view that PT binds a fully extended ACP-bound poly- β -keto intermediate ‘kinked’ into the cyclization chamber, and catalyses formation of the F-folded cyclization pattern (Fig. 4; a rationale for the observed F-folded patterns is detailed in the Supplementary Information and illustrated in Supplementary Fig. 6). In contrast, the ARO/CYC probably bends the extended poly- β -keto intermediate into a hairpin and promotes closure to the S-folded pattern (Supplementary Fig. 6)²⁰. The high-resolution characterizations of a fungal PT and a bacterial ARO/CYC

provide detailed views of these two engines of selective polyketide cyclization and aromatization and will guide the rational alteration of these cyclization domains to new synthetic tasks.

METHODS SUMMARY

All proteins used in the work were overproduced in *E. coli* BL21(DE3) or Rosetta2(DE3) (Novagen), and purified by Ni-affinity followed by FPLC chromatography for crystallization. PT was crystallized at room temperature by sitting-drop vapour diffusion in 0.22 M ammonium acetate (or 0.23 M sodium citrate) and 20–22% PEG3350. Data were collected on beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) and beamline 8.2.2 at the Advanced Light Source (ALS). The PT crystallographic phases were determined by MAD using selenomethionine-derivatized PT. Docking of polyketide substrates and ACP with PT was performed by GOLD²¹ and 3D-Dock²², respectively. CHARMm minimizations were performed with the Discovery Studio (DS) 2.1 suite. PksA global architecture was re-constructed using fragments of PksA detected by native gels and size-exclusion chromatography. Site-directed mutations were performed in the PT monodomain construct pEPT2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 28 August 2009.

1. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
2. Thomas, R. A biosynthetic classification of fungal and *Streptomyces* fused-ring aromatic polyketides. *ChemBioChem* **2**, 612–627 (2001).
3. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **48**, 4688–4716 (2009).
4. Crawford, J. M. *et al.* Deconstruction of iterative multidomain polyketide synthase function. *Science* **320**, 243–246 (2008).
5. Townsend, C. A. & Minto, R. E. in *Comprehensive Natural Products* 443–471 (Elsevier, 1999).
6. Udway, D. W., Merski, M. & Townsend, C. A. Method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J. Mol. Biol.* **323**, 585–598 (2002).
7. Crawford, J. M., Dancy, B. C. R., Hill, E. A., Udway, D. W. & Townsend, C. A. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc. Natl Acad. Sci. USA* **103**, 16728–16733 (2006).
8. Dillon, S. C. & Bateman, A. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinformatics* **5**, 109 (2004).
9. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315–1322 (2008).
10. Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941–953 (2008).
11. Koski, K. M., Haapalainen, A. M., Hiltunen, J. K. & Glumoff, T. Crystal structure of 2-enoyl-CoA hydratase 2 from human peroxisomal multifunctional enzyme type 2. *J. Mol. Biol.* **345**, 1157–1169 (2005).
12. Kimber, M. S. *et al.* The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. *J. Biol. Chem.* **279**, 52593–52602 (2004).
13. Hibbert, F. & Emsley, J. Hydrogen bonding and chemical reactivity. *Adv. Phys. Org. Chem.* **26**, 255–379 (1991).
14. Mills, S. G. & Beak, P. Solvent effects on keto-enol equilibria: tests of quantitative models. *J. Org. Chem.* **50**, 1216–1224 (1985).
15. Jordan, D. B., Zheng, Y. J., Lockett, B. A. & Basarab, G. S. Stereochemistry of the enolization of scytalone by scytalone dehydratase. *Biochemistry* **39**, 2276–2282 (2000).
16. Bahnson, B. J., Anderson, V. E. & Petsko, G. A. Structural mechanism of enoyl-CoA hydratase: three atoms from a single water are added in either an E1cb stepwise or concerted fashion. *Biochemistry* **41**, 2621–2629 (2002).
17. Sachdeva, S. *et al.* Separate entrance and exit portals for ligand traffic in *Mycobacterium tuberculosis* FabH. *Chem. Biol.* **15**, 402–412 (2008).
18. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic *Ascomycetes*. *Proc. Natl Acad. Sci. USA* **100**, 15670–15675 (2003).
19. Ma, S. M. *et al.* Redirecting the cyclization steps of fungal polyketide synthase. *J. Am. Chem. Soc.* **130**, 38–39 (2008).
20. Ames, B. D. *et al.* Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *Proc. Natl Acad. Sci. USA* **105**, 5349–5354 (2008).
21. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).
22. Gabb, H. A., Jackson, R. M. & Sternberg, M. J. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272**, 106–120 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. M. Harris for his gift of HC8. The work at Johns Hopkins was supported by the US National Institutes of Health grant ES001670 awarded to C.A.T. S.-C.T. is supported by the Pew Foundation. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract No. DE-AC02-05CH11231. J.M.C. is currently a fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2002-09, Harvard Medical School).

Author Contributions J.M.C. carried out biochemical experiments and provided recombinant proteins. A.L.V. carried out all mutational studies. T.P.K. assisted by

O.K.-B. determined the PT X-ray crystal structures. E.A.H. prepared substrate and intermediate analogues for co-crystallization experiments. J.W.L. and T.P.K. conducted *in silico* docking studies. J.M.C., T.P.K. and J.W.L. analysed data and contributed to the writing of the paper, and J.W.L. refined the proposed PT mechanism. S.-C.T. and C.A.T. directed the research, provided funding and edited the final manuscript.

Author Information The atomic coordinates of palmitate-bound and HC8-bound PT have been deposited in the Protein Data Bank under accession code 3HRQ and 3HRR. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.A.T. (ctownsend@jhu.edu) or S.-C.T. (sctsai@uci.edu).

METHODS

Protein production and purification. The PT monodomain was amplified by PCR using primers PT5.4 (5'-CGCATATGGAAATCAAGACCACCACGAC-3') and PT3.3n (5'-GTAAGCGGCCGCTAATCGGCCTTTTTCAGTGTAGTCTGC-3'). The amplification product was digested with NdeI/NotI and inserted into the corresponding sites in pET28a (Novagen) to generate pEPT3.3n, a construct encoding the PT domain with a thrombin cleavable N-terminal His₆-tag. The recombinant PksA PT-ACP didomain with a C-terminal His₆-tag fusion was produced and purified as previously described⁴. The PT monodomain was produced in *E. coli* BL21(DE3) (Novagen) by induction with 1 mM IPTG at OD₆₀₀ = 0.6, and incubated at 20 °C overnight. Cell pellets were lysed by sonication in 50 mM potassium phosphate (pH 7.6), 10 mM imidazole, 300 mM NaCl, and 10% glycerol. The His₆-tagged protein was purified by Qiagen Ni-NTA resin, cleaved with bovine thrombin (10 units per 1 mg sample) for 48 h at 16 °C to remove the His₆-tag, dialysed against 50 mM Tris-HCl pH 7.5, and further purified over a Q-Sepharose anion exchange column (10–500 mM KCl gradient). The resulting pure protein was dialysed against 20 mM Tris-HCl (pH 7.5), 5% glycerol, and 2 mM dithiothreitol for crystallographic studies.

Se-Met-labelled protein production. pEPT3.3n was transformed into a methionine auxotrophic strain, B834(DE3) (Novagen). Overnight cultures were carried out in standard LB medium with 25 µg ml⁻¹ kanamycin. The cells were harvested, washed once and added to 200 ml of LeMaster's medium containing 25 µg ml⁻¹ kanamycin. The culture was grown at 37 °C to exchange the Met pool with Se-Met. Aliquots were used to inoculate 4–1 L LeMaster's cultures. The cultures were grown to an OD₆₀₀ = 0.6–0.8 at 37 °C and induced with 1 mM IPTG. The induction temperature was reduced to 20 °C, and cultivated for an additional 16–18 h. Protein purifications were carried out as described for non-labelled proteins.

Crystallization, structure determination and refinement. Crystals of the selenomethionine derivative of PksA PT-ACP construct were grown by vapour diffusion in sitting drops over 500 µl of the well solution at 25 °C. Drops were generated by mixing 2 µl purified protein at 10 mg ml⁻¹ with 2 µl well solution (0.23 M sodium citrate, 22% PEG 3350). Analysis of PT-ACP crystals by SDS-PAGE showed that it degraded into the PT and ACP monodomains, and the PT domain selectively crystallized as a protein degradation product. Tetragonal crystals grew overnight. Orthorhombic crystals of the PT monodomain construct (different from the degradation product) grew in a similar condition (0.22 M ammonium acetate, 20% PEG3350) overnight at 25 °C. Crystals were soaked in crystallization solution plus 30% glycerol and flash frozen in liquid nitrogen before data collection. The HC8-bound PksA PT crystals were obtained by treating the monodomain PT with Amberlite XAD-2 resin for 2 h, followed by addition of HC8 (500 µM in acetone) in a 10:1 (HC8:PT) molar ratio. Orange crystals grew overnight in the same condition as the monodomain PT. Data were

collected on beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL), and subsequently integrated, indexed and scaled with HKL2000. Phases were determined by MAD using ShelX²³ and Solve/Resolve²⁴. All 20 Se sites except for the two N-terminal methionines were found. Electron density maps were generated using CNS and the model was built using COOT²⁵. Refinement of the SeMet PT model was carried out with CNS²⁶. The native palmitate-bound PT and HC8-bound PT crystal structures were solved by molecular replacement using the SeMet PT model as the template, using CNS (native PT) and CCP4 (HC8-PT)²⁷. Waters were added in COOT followed by refinement with CNS or CCP4, and edited by visual inspection. Initial coordinates and restraints for palmitate and HC8 were generated with PRODRG²⁸ followed by manual fitting into well-defined positive $F_o - F_c$ electron density contoured at 4σ with COOT. Supplementary Table 1 lists a summary of phasing and refinement statistics. The refined structures were evaluated with PROCHECK²⁹ before coordinate submission.

Mutagenesis, *in vitro* PksA PT assay and HPLC product analysis. Three-part multidomain combination reactions (SAT-KS-MAT + PT* + *holo*-ACP) and subsequent product analysis by HPLC were carried out as previously described⁴. Site-directed mutations (*) in the PT monodomain construct (pEPT2) were performed using overlap-extension PCR³⁰.

Native PAGE analysis. The enzymes were dialysed against 10 mM Tris-HCl pH 7.5, 10% glycerol and 2 mM dithiothreitol to avoid negative salt effects on native PAGE. The proteins were compared to non-denatured protein from the Sigma-Aldrich molecular weight marker kit according to the manufacturer's instructions. Association states: SAT, monomer; SAT-KS-MAT, dimer; PT, dimer; ACP, monomer; TE, monomer; PT-ACP, dimer; ACP-TE, monomer; PT-ACP-TE, dimer/monomer (Supplementary Table 2).

23. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
24. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
25. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
26. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
27. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
30. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).

LETTERS

Structural basis for translational fidelity ensured by transfer RNA lysidine synthetase

Kotaro Nakanishi^{1†}, Luc Bonnefond², Satoshi Kimura³, Tsutomu Suzuki³, Ryuichiro Ishitani² & Osamu Nureki^{1,2}

Maturation of precursor transfer RNA (pre-tRNA) includes excision of the 5' leader and 3' trailer sequences, removal of introns and addition of the CCA terminus^{1–3}. Nucleotide modifications are incorporated at different stages of tRNA processing, after the RNA molecule adopts the proper conformation. In bacteria, tRNA^{Ile2} lysidine synthetase (TilS) modifies cytidine into lysidine (L; 2-lysyl-cytidine) at the first anticodon of tRNA^{Ile2} (refs 4–9). This modification switches tRNA^{Ile2} from a methionine-specific to an isoleucine-specific tRNA^{Ile2}. However, the aminoacylation of tRNA^{Ile2} by methionyl-tRNA synthetase (MetRS), before the modification by TilS, might lead to the misincorporation of methionine in response to isoleucine codons. The mechanism used by bacteria to avoid this pitfall is unknown. Here we show that the TilS enzyme specifically recognizes and modifies tRNA^{Ile2} in its precursor form, thereby avoiding translation errors. We identified the lysidine modification in pre-tRNA^{Ile2} isolated from RNase-E-deficient *Escherichia coli* and did not detect mature tRNA^{Ile2} lacking this modification. Our kinetic analyses revealed that TilS can modify both types of RNA molecule with comparable efficiencies. X-ray crystallography and mutational analyses revealed that TilS specifically recognizes the entire L-shape structure in pre-tRNA^{Ile2} through extensive interactions coupled with sequential domain movements. Our results demonstrate how TilS prevents the recognition of tRNA^{Ile2} by MetRS and achieves high specificity for its substrate. These two key points form the basis for maintaining the fidelity of isoleucine codon translation in bacteria. Our findings also provide a rationale for the necessity of incorporating specific modifications at the precursor level during tRNA biogenesis.

tRNA biosynthesis is a complex, multi-step process that requires numerous protein factors^{1–3}. Post-transcriptional modifications are incorporated during all stages of pre-tRNA processing. For *E. coli* tRNAs, modifications such as T, Ψ , D and m²A occur at early stages, whereas m¹G and Gm occur at later stages¹⁰. In all cases, the precursor tRNA conformation is critical for the recognition by modification enzymes^{11,12}. The lysidine modification is introduced at the wobble position of the bacterial tRNA^{Ile2} anticodon (C34) by TilS enzyme^{4–9}. This modification concomitantly switches the amino-acid coding and aminoacyl-tRNA synthetase (aaRS) specificity from methionine to isoleucine *in vitro*^{8,9,13}. Previous results, however, have stressed the potential risk of modifying mis-acylated Met-tRNA^{Ile2} for translational fidelity (Supplementary Fig. 1).

To address this issue, we first determined the stage of tRNA^{Ile2} processing at which TilS introduces lysidine. We isolated pre- and mature tRNA^{Ile2} from a temperature-sensitive mutant of *E. coli* RNase E and determined their post-transcriptional modifications by mass spectrometry^{14–17} (Fig. 1a). Assignment of the fragments revealed that pre-tRNA^{Ile2} lacks a 5'-leader sequence, but has a 39–40 nucleotide

3'-trailer sequence (Fig. 1b). The anticodon-containing fragments of both pre- and mature tRNA^{Ile2} include lysidine at the wobble position (Fig. 1c and Supplementary Fig. 2). However, whereas all of the observed mature tRNA^{Ile2} fragments contained only L34, we detected a small amount of pre-tRNA^{Ile2} fragments with unmodified C34. This result suggests that tRNA^{Ile2} is modified by TilS before and during processing by RNase E, before its full maturation. This property excludes the possibility of generating Met-tRNA^{Ile2}_{CAU} and its accompanying misdecoding (Supplementary Fig. 1). We also detected other post-transcriptional modifications. s⁴U8 and t⁶A37 are present in both pre- and mature tRNA^{Ile2}, whereas Gm18 is only found in the mature tRNA (Fig. 1b and Supplementary Table 1). These discrepancies confirm that the recognition of pre-tRNAs is not a general property of modification enzymes, but is a critical adaptation of TilS to ensure the fidelity of isoleucine codon translation in bacteria (Supplementary Fig. 1). To characterize TilS activity on pre-tRNA^{Ile2} more precisely, we determined the kinetic parameters for different tRNA transcripts containing extra sequences. The measured parameters were roughly similar for all of the tested molecules, including mature tRNA (wild type), 5'- and/or 3'-containing precursors (5L-tRNA^{Ile2}, tRNA^{Ile2}-6T and 5L-tRNA^{Ile2}-6T) and methionylated tRNA (Met-tRNA^{Ile2}) (Table 1). This uniform binding suggests that TilS precisely recognizes the tRNA structure in the precursor, without the participation of the untrimmed sequences. Thus, TilS can modify pre-tRNA^{Ile2} in the early stages of the processing pathway, as soon as the proper tRNA scaffold is accessible.

The second key point of the TilS activity is its specificity for the RNA substrate. Although TilS activity resembles that of aaRS, the lysine group is covalently conjugated to C34 by TilS and cannot be hydrolysed under physiological conditions⁶, unlike the labile ester bond of aaRS-generated aminoacyl-tRNAs. Lysidinylation of non-cognate (pre-)tRNAs would disturb the quantity balance of mature tRNAs for protein synthesis. To understand how TilS discriminates pre-tRNA^{Ile2} from other pre- or mature tRNAs, we solved the crystal structure of *Geobacillus kaustophilus* TilS (GkTilS) complexed with *Bacillus subtilis* tRNA^{Ile2}_{CAU} at 3.65-Å resolution, by the multi-wavelength anomalous dispersion method (Fig. 2a, Supplementary Fig. 3 and Supplementary Table 2). The asymmetric unit contains one TilS homodimer and two tRNAs, each tightly embedded in one monomer of TilS, with an overall interface of 2,998 Å². Each monomer consists of an amino-terminal catalytic domain, and two carboxy-terminal domains, connected by a long α -helical linker and a loop linker, respectively. GkTilS, like *E. coli* TilS (EcTilS), is a type-I TilS enzyme, as defined by the presence of the second C-terminal domain. Unexpectedly, the structure of the tRNA-bound GkTilS revealed large movements of the two C-terminal domains compared with the apo-EcTilS (Fig. 3a and Supplementary Discussion). The

¹Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Kanagawa 225-8501, Japan. ²Department of Basic Medical Sciences, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ³Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. †Present address: Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.

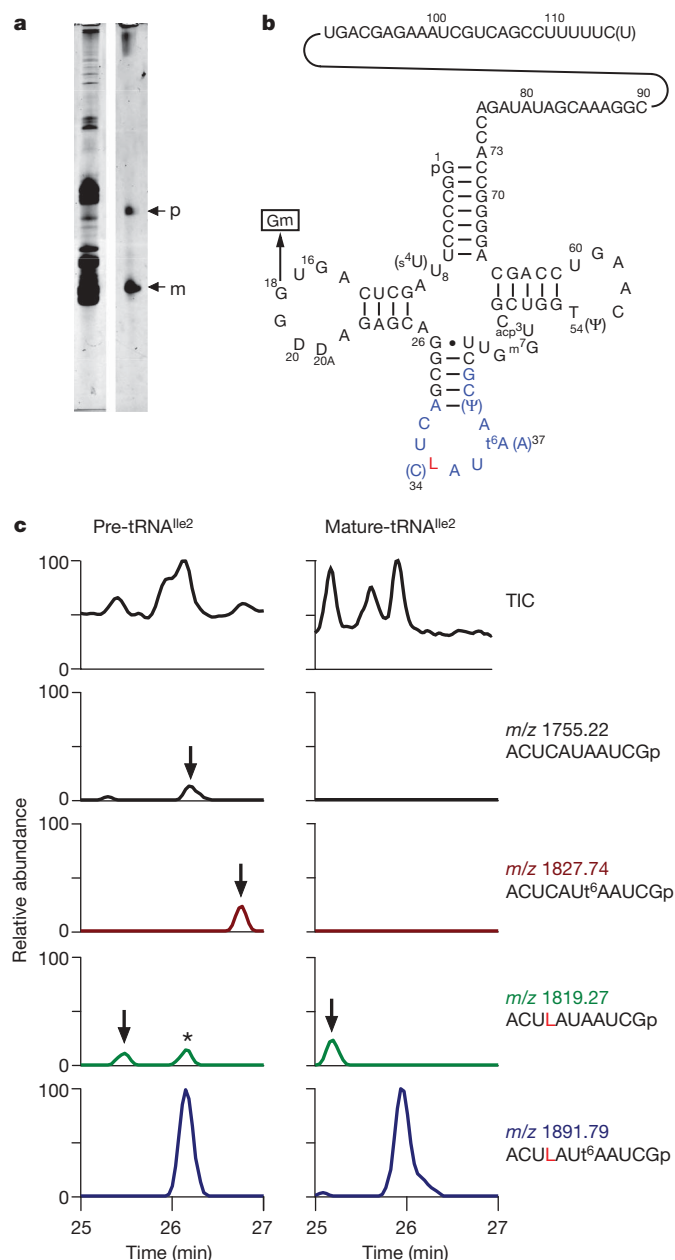


Figure 1 | Mass spectrometric analysis of pre- and mature tRNA^{Ile2}s purified from *E. coli*. **a**, Purified tRNA^{Ile2}s from the *E. coli* RNase E mutant (right), showing precursor (p) and mature (m) forms. The left lane represents total RNA. **b**, Purified *E. coli* pre-tRNA^{Ile2} bearing a 3'-trailer sequence with modified nucleosides: lysidine (L), 4-thiouridine (s⁴U), N⁶-threonylcarbamoyladenosine (t⁶A), pseudouridine (Ψ), 7-methylguanosine (m⁷G), 3-(3-amino-3-carboxypropyl)uridine (acp³U) and 5-methyluridine (T). **c**, Liquid chromatography/mass spectrometry analyses of precursor (left) and mature (right) forms of tRNA^{Ile2} digested by RNase T₁. The peaks of the anticodon-containing fragments are shown by arrows. The peak with an asterisk is derived from a different fragment.

first C-terminal domain interacts with the inner corner of the L-shaped tRNA, as seen for the stem contact fold of the class Ia/Ib aaRS (Supplementary Fig. 4)^{18,19}. Simultaneously, the second C-terminal domain binds the acceptor stem of tRNA^{Ile2} from the major groove side (Fig. 2a). The additional leader and trailer sequences of pre-tRNA^{Ile2} can be accommodated on this domain without steric clashes. On the basis of these observations, we hereafter refer to the first and second C-terminal domains as the stem contact-like (SCL) and acceptor stem-binding (ASB) domains, respectively. The extensive recognition of the entire tRNA scaffold by the type-I

Table 1 | Kinetic parameters of GktRNA^{Ile2} transcript lysidinylation by GkTilS

tRNA ^{Ile2}	K _m (μM)	10 ⁻³ k _{cat} (s ⁻¹)	k _{cat} /K _m (s ⁻¹ μM ⁻¹)	Relative k _{cat} /K _m
Mature tRNA				
Wild type	0.14 ± 0.03	30 ± 15	0.214	1
Precursor tRNA				
5L-tRNA	0.08 ± 0.01	25 ± 1	0.312	1.46
tRNA-6T	0.13 ± 0.03	40 ± 5	0.308	1.44
5L-tRNA-6T	0.18 ± 0.06	40 ± 5	0.222	1.02
Aminoacylated tRNA				
Met-tRNA	0.05 ± 0.01	30 ± 5	0.6	2.8

Values are mean ± s.e.m.

TilS is unlike those of other anticodon-modifying enzymes, such as QueTGT, TadaA and MnMA, which only recognize the anticodon arm^{20–22}.

The tRNA anticodon arm interacts with the SCL and catalytic domains, and the anticodon loop undergoes a substantial conformational change (Fig. 2b and Supplementary Figs 5 and 12). The mismatched base pair C32:A38, which is conserved in the canonical tRNA^{Phe} structure, is disrupted, with A38 flipped out^{23,24}. In this conformation, A38 can be accommodated in a hydrophobic pocket of the SCL domain. Indeed, the replacement of A38 with C in tRNA^{Ile2} drastically reduced the lysine-accepting activity (Fig. 2d), consistent with a previous mutational analysis of the hydrophobic pocket²⁵. Our structure also showed that Arg 142 replaces the flipped-out A38 and mimics the base pairing with the Watson–Crick face of C32. This idea is supported by the dramatic loss of TilS activity when either GkTilS Arg 142 is substituted by Ala or tRNA^{Ile2} C32 by U (Fig. 2d). This double-lock system tightly anchors the tRNA anticodon loop and places the target C34 base at the entrance of the catalytic tunnel, to facilitate the lysidinylation reaction (Fig. 2b). A similar mechanism was observed in the structure of the pseudouridine synthetase RluA complexed with the tRNA^{Phe} anticodon arm²⁶. In this tunnel, the strictly conserved residues Asp 131, Glu 134 and Arg 199, which surround the target C34, are likely to be involved in the activation of C34 with ATP and the subsequent lysidinylation reaction, as suggested by the previous mutagenic analyses (Supplementary Figs 5 and 6)²⁷.

At the other extremity, the tRNA molecule is recognized by the ASB domain. This domain, whose structure was solved independently at 2.2-Å resolution, consists of two antiparallel β-sheets with an inserted β-hairpin and a subsequent helix–turn–helix (HTH) motif, forming a pseudo-knot structure (Supplementary Fig. 7 and Supplementary Table 3). The β-hairpin stacks against the tRNA 3'-ACCA terminus and the G1:C72 base pair in a shape-complementary manner. These observations are consistent with the limited effect of their base substitution on the modification activity (Fig. 2d). The binding mode of the HTH motif to the tRNA major groove is reminiscent of that from transcriptional regulatory proteins (Supplementary Discussion)²⁸. The first α-helix of the HTH motif, including bulkier aliphatic residues (Lys 404, Lys 405 and Lys 407), is snugly accommodated in the major groove of the acceptor stem, near the C4:G69 and C5:G68 base pairs, which are known as identity determinants for *Ec*TilS (Fig. 2c)²⁵. However, the GkTilS Lys404Ala and Lys405Ala mutants only exhibited moderately decreased lysidine-modification abilities, and the Lys407Ala, Glu408Ala and Glu412Ala mutants retained almost the same level of activity as the wild-type enzyme (Fig. 2d). These discrepancies suggest that the interaction between the ASB domain and the acceptor stem is too tight to be disrupted by a single amino-acid substitution. Indeed, our gel-shift analysis showed that the ASB domain is able to bind tRNA^{Ile2} by itself (Supplementary Fig. 7).

A comparison of the complex structure with that of the *Ec*TilS apo-form (PDB accession 1NI5) revealed large movements of the ASB and SCL domains (Supplementary Discussion). The highly conserved Arg 420 and Phe 410, on the HTH motif, form a hydrophobic

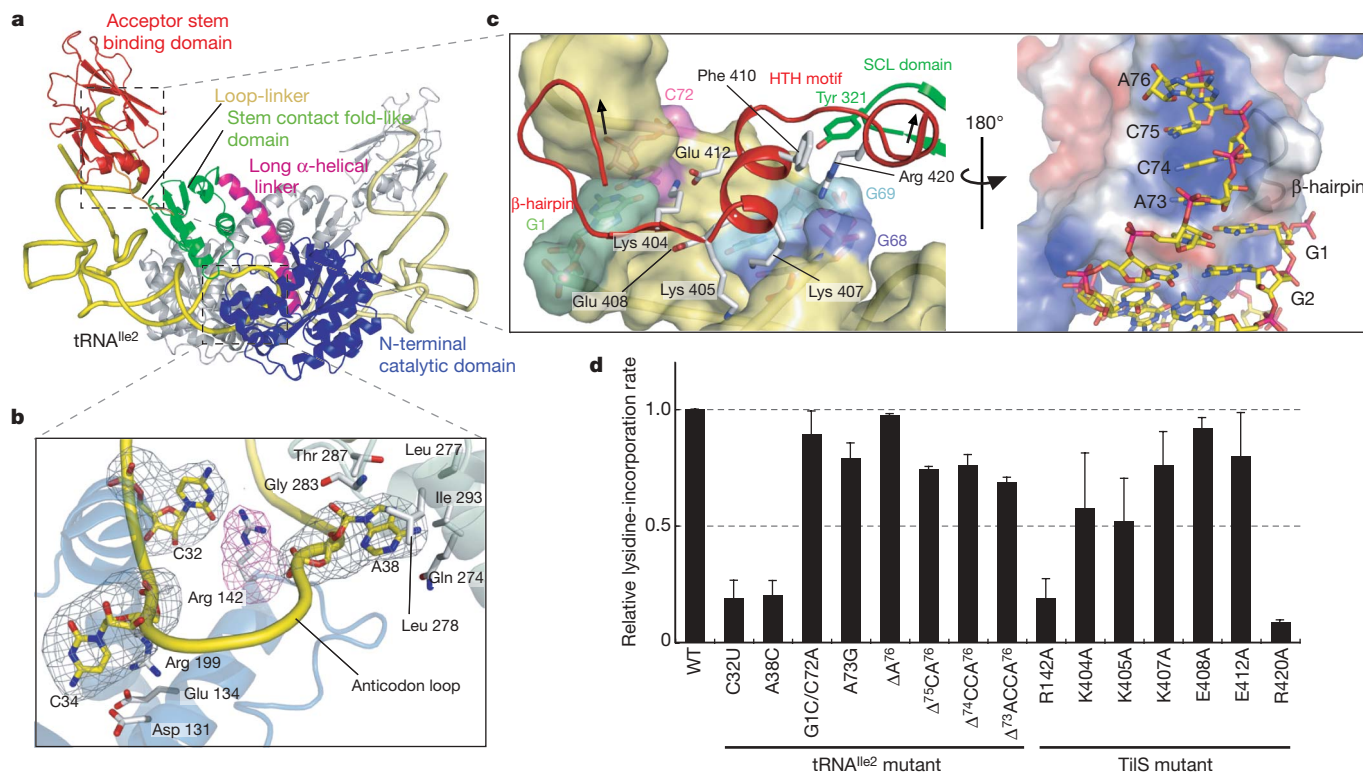


Figure 2 | tRNA recognition by GkTilS. **a**, Overall homodimeric structure. **b**, Recognition of the anticodon loop. $F_o - F_c$ simulated annealing omit maps (3.7σ) of Arg 142 and three nucleotides (C32, C34 and A38) are shown in magenta and grey, respectively. **c**, Left, recognition of the major groove of the acceptor stem by the HTH motif and β -hairpin; right, the protein surface

that complementarily interacts with the 3'-ACCA terminus is coloured according to its electrostatic potential. **d**, Lysine incorporating activities of tRNA^{Ile2} and GkTilS mutants. The initial rate of lysine incorporation is shown. Error bars, s.d. of three independent experiments.

core with Tyr 321, on the SCL domain, in both the complex and apo-forms (Fig. 2c). However, the other hydrophobic interactions between the HTH and the SCL domain are disrupted upon tRNA binding, and are replaced by electrostatic interaction between the HTH motif and the tRNA acceptor helix (Supplementary Fig. 8). Consequently, the ASB domain rotates outward by 59.4° relative to the SCL domain (Fig. 3b). Indeed, the Arg420Ala mutant displayed significantly decreased activity (Fig. 2d), supporting the idea that the above interactions might be essential to lock the rearranged conformation. To examine the non-biased dynamics of tRNA-free TilS, we conducted a 10-ns molecular-dynamics simulation of the apo-EcTilS. The interactions between the HTH and the SCL domain were retained during the simulation (Supplementary Fig. 9), implying that they are stably formed in solution. Furthermore, the direction of the largest fluctuation mode observed in this simulation is consistent with that of the present domain movements from the apo- to the tRNA-bound form (Supplementary Fig. 9). This intrinsic fluctuation of the apo-form may be amplified upon tRNA acceptor-end capture by the ASB domain, resulting in the drastic movement of the ASB and SCL domains (Fig. 3a, b). In the apo-form, the SCL domain interacts with both of the kinked long α -helical linkers to place itself close to the N-terminal catalytic domain, where it covers the entrance of the A38-binding pocket (Fig. 3b, e). In contrast, the α -helical linker is straightened in the complex form, and the SCL domain rotates outward by 35.3° relative to the catalytic domain, so that the A38-binding pocket can grasp the edge of the anticodon loop (Fig. 3b, g). Whereas the exposure of this pocket requires the tRNA-induced structural change, the acceptor-end binding surface on the ASB domain is already formed in the apo-form structure (Fig. 3f). These contrasting observations suggest that the capture of the acceptor end by the ASB domain and the subsequent structural changes precede the binding of A38 by the SCL domain. This view is supported by our gel-shift

analyses, in which the ASB domain bound tRNA^{Ile2} by itself, whereas its deletion abolished both the tRNA binding and enzymatic activities of TilS (Supplementary Fig. 7)²⁵. On the basis of these observations, we propose the following sequential tRNA recognition mechanism (Fig. 3d–g): the ASB domain first captures the acceptor arm; the ASB and SCL domains then rotate while grasping the tRNA; finally, the anticodon loop is precisely placed in the 'loop-binding crater'. Therefore, the interaction with the tRNA starts at the acceptor stem furthest from the target C34, and TilS then recognizes the tRNA^{Ile2} identity determinants, G69, G68, A38 and C32, step-by-step towards C34. No other tRNA in *G. kaustophilus* has this set of nucleotides (Supplementary Discussion). The tRNA recognition accompanied by these sequential structural transitions may confer an allosteric aspect to the tRNA binding kinetics of the type-I TilS, thereby increasing the specificity for the cognate pre-tRNA^{Ile2} over the non-cognate (pre-)tRNAs. To our knowledge, such a drastic domain reorganization has not been observed in any other tRNA recognition enzymes (Supplementary Discussion). The tRNA modification enzymes generally have low values of Michaelis constant (K_m) for tRNA compared with aaRS (Supplementary Table 4). The higher specificity of the modification enzymes for their (pre-)tRNAs might compensate for the absence of a quality-control system, such as the aaRS editing mechanisms. The high K_m of type-II *Aquifex aeolicus* TilS (AaTilS) lacking the ASB domain for tRNA is consistent with its moderate mis-modification of the near-cognate tRNA^{Met} (Fig. 3c). Interestingly, apo-AaTilS already has the same conformation as tRNA-bound GkTilS (Fig. 3b), suggesting that AaTilS recognizes tRNA^{Ile2} without any conformational change. The present sequential and time-consuming RNA recognition mechanism provides high accuracy in tRNA^{Ile2} recognition (Supplementary Table 4), as shown by the ability of TilS to discriminate tRNA^{Ile2} from two tRNA^{Met} isoacceptors (Fig. 3c). Such an enhanced specificity for tRNA^{Ile2} recognition, which cannot

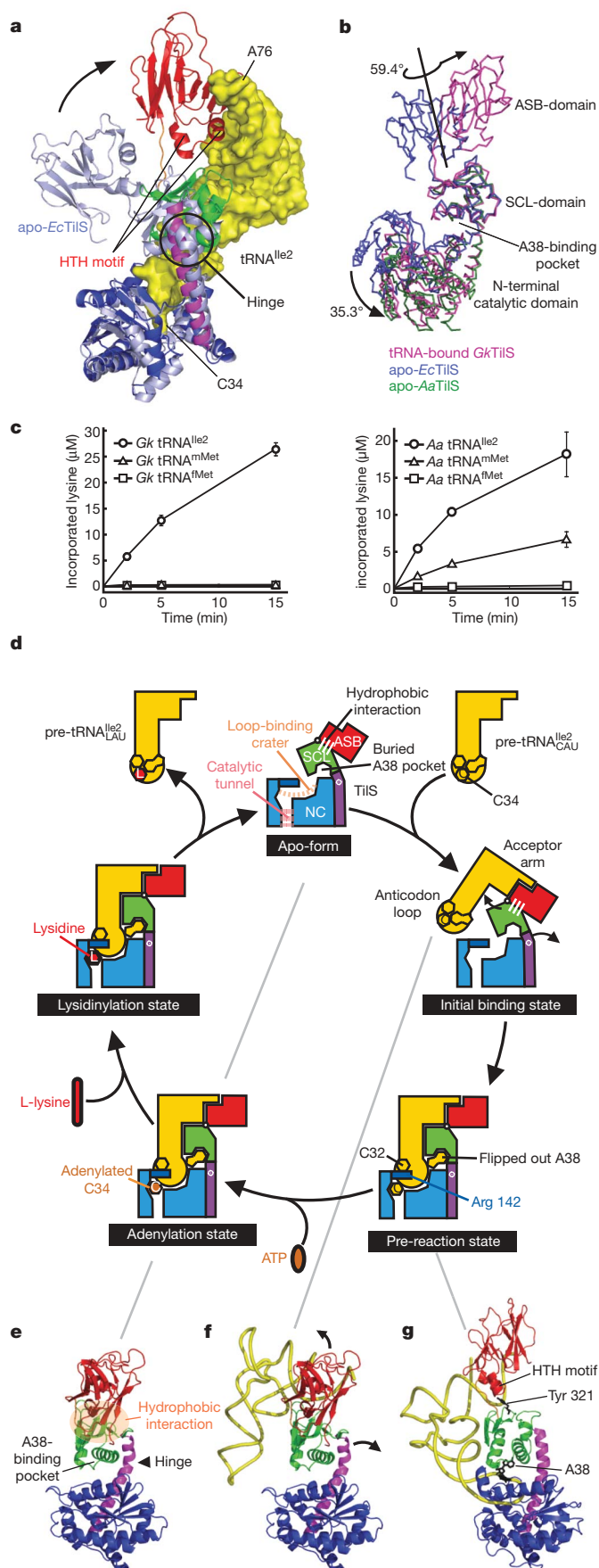


Figure 3 | Sequential tRNA recognition mechanism. **a**, Superimposition of apo-EcTilS and tRNA-bound GkTilS on their catalytic domains.

b, Superimposition of type I and II TilSs on their SCL domains. **c**, Lysine incorporation into tRNA^{Met}s by GkTilS (left) and AaTilS (right). Error bars denote s.d. of three independent experiments. **d**, Lysidine formation model. **e**, Apo-form (apo-EcTilS): the ASB and SCL domains have an intramolecular hydrophobic interaction. **f**, Initial binding state (manual docking of yeast tRNA^{Phe} (PDB accession 1EHZ) onto the apo-EcTilS): capture of the pre-tRNA^{Ile2} acceptor stem by the ASB domain triggers disruption of the hydrophobic interactions. **g**, Pre-reaction state (the current structure): drastic domain movements allow TilS to interact fully with tRNA.

METHODS SUMMARY

About 100 fmol of the pre-tRNA^{Ile2} or mature tRNA^{Ile2} (purified as described in the Methods) was digested with RNase T₁ or RNase A, and the digests analysed by capillary liquid-chromatography nano-electrospray ionization/mass spectrometry¹⁷. We used a linear ion trap-orbitrap hybrid mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific), equipped with a custom-made nanospray ion source, and a splitless nano-high-performance liquid chromatography system (DiNa, KYA Technologies).

B. subtilis tRNA^{Ile2} (BstRNA^{Ile2}) was prepared as reported previously²⁷. The only sequence difference between *G. kaustophilus* tRNA^{Ile2} and BstRNA^{Ile2} is the base at position 16, which GkTilS did not contact in the crystal structure. Native crystals of the GkTilS-BstRNA^{Ile2} complex were obtained by the vapour diffusion method, in 45 mM Na-cacodylate, pH 6.5, 200 mM NaCl, 225 mM ammonium acetate, 11 mM CaCl₂ and 7.2% PEG 4,000. Data collection and refinement statistics are shown in Supplementary Table 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 May; accepted 2 September 2009.

1. Bjork, G. R. in *tRNA: Structure, Biosynthesis and Function* (eds Söll, D. & RajBhandary, U. L.) 165–206 (ASM Press, 1995).
2. Hendrickson, T. & Schimmel, P. in *Translation Mechanisms* (eds Lapointe, J. & Brakier-Gingras, L.) 33–64 (Landes Bioscience, 2003).
3. Nakanishi, K. & Nureki, O. Recent progress of structural biology of tRNA processing and modification. *Mol. Cells* **19**, 157–166 (2005).
4. Harada, F. & Nishimura, S. Purification and characterization of AUA specific isoleucine transfer ribonucleic acid from *Escherichia coli* B. *Biochemistry* **13**, 300–307 (1974).
5. Matsugi, J., Murao, K. & Ishikura, H. Characterization of a *B. subtilis* minor isoleucine tRNA deduced from tDNA having a methionine anticodon CAT. *J. Biochem.* **119**, 811–816 (1996).
6. Muramatsu, T. *et al.* A novel lysine-substituted nucleoside in the 1st position of the anticodon of minor isoleucine transfer-RNA from *Escherichia coli*. *J. Biol. Chem.* **263**, 9261–9267 (1988).
7. Weber, F., Dietrich, A., Weil, J. H. & Marechal-Drouard, L. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucleic Acids Res.* **18**, 5027–5030 (1990).
8. Soma, A. *et al.* An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol. Cell* **12**, 689–698 (2003).
9. Muramatsu, T. *et al.* Codon and amino-acid specificities of a transfer-RNA are both converted by a single post-transcriptional modification. *Nature* **336**, 179–181 (1988).
10. Sakano, H. & Shimura, Y. Characterization and *in vitro* processing of transfer RNA precursors accumulated in a temperature sensitive mutant of *Escherichia coli*. *J. Mol. Biol.* **123**, 287–326 (1978).
11. Etcheverry, T., Colby, D. & Guthrie, C. Precursor to a minor species of yeast tRNA^{Ser} contains an intervening sequence. *Cell* **18**, 11–26 (1979).
12. Nishikura, K. & De Robertis, E. M. RNA processing in micro injected *Xenopus laevis* oocytes. Sequential addition of base modifications in a spliced transfer RNA. *J. Mol. Biol.* **145**, 405–420 (1981).
13. Nakanishi, K., Ogiso, Y., Nakama, T., Fukai, S. & Nureki, O. Structural basis for anticodon recognition by methionyl-tRNA synthetase. *Nature Struct. Mol. Biol.* **12**, 931–932 (2005).
14. Kimata, K., Tanaka, Y., Inada, T. & Aiba, H. Expression of the glucose transporter gene, ptsG, is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J.* **20**, 3587–3595 (2001).
15. Kaneko, T. *et al.* Wobble modification differences and subcellular localization of tRNAs in *Leishmania tarentolae*: implication for tRNA sorting mechanism. *EMBO J.* **22**, 657–667 (2003).
16. Suzuki, T. & Suzuki, T. Chaplet column chromatography: isolation of a large set of individual RNAs in a single step. *Methods Enzymol.* **425**, 231–239 (2007).
17. Suzuki, T., Ikeuchi, Y., Noma, A., Suzuki, T. & Sakaguchi, Y. Mass spectrometric identification and characterization of RNA-modifying enzymes. *Methods Enzymol.* **425**, 212–229 (2007).

be achieved by a rigid enzyme, may be required for the sophisticated switching of the tRNA^{Ile2} specificities for both the codon and amino acid by the type-I TilSs, to maintain the fidelity of the genetic code.

18. Sugiura, I. *et al.* The 2.0 Å crystal structure of *Thermus thermophilus* methionyl-tRNA synthetase reveals two RNA-binding modules. *Structure* **8**, 197–208 (2000).
19. Rould, M. A., Perona, J. J. & Steitz, T. A. Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase. *Nature* **352**, 213–218 (1991).
20. Numata, T., Ikeuchi, Y., Fukai, S., Suzuki, T. & Nureki, O. Snapshots of tRNA sulphuration via an adenylated intermediate. *Nature* **442**, 419–424 (2006).
21. Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nature Struct. Mol. Biol.* **13**, 153–159 (2006).
22. Xie, W., Liu, X. J. & Huang, R. H. Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nature Struct. Biol.* **10**, 781–788 (2003).
23. Nagaswamy, U., Voss, N., Zhang, Z. D. & Fox, G. E. Database of non-canonical base pairs found in known RNA structures. *Nucleic Acids Res.* **28**, 375–376 (2000).
24. Shi, H. J. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105 (2000).
25. Ikeuchi, Y. *et al.* Molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition. *Mol. Cell* **19**, 235–246 (2005).
26. Hoang, C. *et al.* Crystal structure of pseudouridine synthase RluA: Indirect sequence readout through protein-induced RNA structure. *Mol. Cell* **24**, 535–545 (2006).
27. Nakanishi, K. *et al.* Structural basis for lysidine formation by ATP pyrophosphatase accompanied by a lysine-specific loop and a tRNA-recognition domain. *Proc. Natl Acad. Sci. USA* **102**, 7487–7492 (2005).
28. Jordan, S. R. & Pabo, C. O. Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. *Science* **242**, 893–899 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the beamline staff at BL41XU of SPring-8 (Harima, Japan) and NW12 of PF-AR (Tsukuba, Japan) for help during data collection. We thank H. Aiba for providing the temperature-sensitive *E. coli* strain of RNase E, A. Soma and Y. Sekine for discussions, and T. Suzuki and Y. Sakaguchi for support with mass spectrometry. This work was supported by a SORST program grant from Japan Science and Technology to O.N., by a grant for the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N. and T.S., by grants from MEXT to R.I. and O.N., and by Mitsubishi Foundation and Kurata Memorial Hitachi Science and Technology Foundation grants to O.N. L.B. and K.N. were supported by a postdoctoral fellowship for foreign researchers and a young scientist fellowship, respectively, from the Japan Society for the Promotion of Science.

Author Contributions K.N. performed purification, crystallization and structure determination of the complex, and biochemical analyses. L.B. performed purification, crystallization and structure determination of the ASB domain, and biochemical analyses. S.K. purified the native tRNAs and performed the mass spectrometry analysis under the supervision of T.S. R.I. performed the molecular dynamics analysis and assisted with the structural determination. O.N. assisted with the structure determination. All authors discussed the results and commented on the manuscript. O.N. supervised all the work.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 3A2K (GkTilS–tRNA complex) and 3HJ7 (ASB domain). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.N. (nureki@ims.u-tokyo.ac.jp).

METHODS

Isolation of *E. coli* pre-tRNA^{Ile2}. The temperature-sensitive *E. coli* strain of RNase E (GW20, W3110 *ams-1 zce726::Tn10*)¹⁴ was provided by H. Aiba. GW20 cells were cultivated at 30 °C. When the A₆₀₀ reached 0.7, the temperature was raised to 43 °C. After 30 min of cultivation, the cells were collected. Total RNAs (200 A₂₆₀ units) were extracted from the cells using the guanidinium thiocyanate/phenol/chloroform method²⁹. Accumulation of pre-tRNA^{Ile2} was confirmed by northern blotting⁸. Precursor and mature forms of tRNA^{Ile2} were purified from the total RNA by the solid-phase DNA probe method described previously^{15,16}. The 3'-biotinylated synthetic DNA oligonucleotide, 5'-TATA AGTCGCTGCTCTAACCCTGAGCTA-3', was used as a probe. Both the pre- and mature tRNA^{Ile2}s were further purified by PAGE.

Protein purification. The gene encoding *G. kaustophilus* TlS was cloned into the pET28c plasmid and overexpressed in *E. coli* BL21 (DE3)-RIL. The cells expressing the recombinant protein were lysed by sonication in 100 mM Na-phosphate, pH 6.6, 100 mM NaCl and 1 mM phenylmethylsulphonyl fluoride (PMSF), then centrifuged. After incubation of the supernatant at 47 °C for 30 min followed by centrifugation, the GkTlS protein was purified by SP-Sepharose, Resource PHE and Heparin affinity chromatography (GE Healthcare). The eluted sample was dialysed against 20 mM MES, pH 6.0, 300 mM NaCl, 100 mM ammonium sulphate and 1 mM dithiothreitol (DTT) at 20 °C for 12 h. The purified TlS was concentrated to about 6 mg ml⁻¹ and was used for crystallization. Point mutations within pET28c, encoding C-terminally His-tagged GkTlS, were introduced with a Quikchange Site-Directed Mutagenesis kit (Stratagene). The gene encoding GkTlS ΔASB was also cloned into pET28c. His-tagged wild-type GkTlS and its mutants were purified using Ni-NTA (Qiagen) and Heparin affinity columns (GE Healthcare). The eluted samples were dialysed against 50 mM Tris, pH 8.0, 400 mM NaCl, 10 mM β-mercaptoethanol and 50% glycerol at 20 °C for 12 h before storage at -20 °C. The DNA encoding the GkTlS ASB domain was cloned into the pET22b plasmid and overexpressed. The protein was purified in the same way as the full-length GkTlS, except for the use of a HiLoad 16/60 Superdex200 column (GE Healthcare) after the Resource PHE column. The eluted sample was dialysed against 20 mM MES, pH 6.5, 400 mM NaCl and 5 mM DTT at 20 °C for 12 h, concentrated to about 9.5 mg ml⁻¹ and used for crystallization. AaTlS was purified as described previously²⁷.

Structure determination and refinement. Because TlS only recognizes the tRNA structure within the precursor molecule, we used the mature tRNA^{Ile2} transcript, to avoid the negative effect of the unbound extra-sequences on crystallization. Unfortunately, the *E. coli* complex crystals did not diffract beyond 7-Å resolution. We therefore switched to the *Geobacillus kaustophilus* complex. Selenomethionine (SeMet)-substituted GkTlS was purified in the same manner as the wild-type enzyme. *G. kaustophilus* tRNA^{Ile2} (GkRNA^{Ile2}) was prepared as reported previously²⁷. The SeMet-substituted crystal of GkTlS-GkRNA^{Ile2} was soaked in the collection buffer (1.2-fold concentrated reservoir solution), cryoprotected with 30% glycerol and flash-cooled in a nitrogen stream at 100 K. The SeMet derivative data set was collected at BL41XU of SPring-8, using an ADSC Q315 CCD (charge-coupled device) detector, and was processed using the program SnB³¹, and were used for phase calculation at 4.0-Å resolution with the program SHARP³². The initial phases were improved by solvent flattening with the program SOLOMON³³ and by non-crystallographic symmetry averaging with the program CCP4 (ref. 34). The initial model was manually built with the program O³⁵, and was improved by iterative cycles of refinement with the programs Phenix³⁶ and Refmac³⁷. In the tracing of the main chain, the crystal structures of the TlSs from *E. coli* and *A. aeolicus* were used as guide structures. Native data of GkTlS complexed with BstRNA^{Ile2} were collected at NW-12 of KEK, using an ADSC-Q210 CCD detector. Molecular replacement was performed with the program MOLREP³⁸, using the SeMet structure as a search model. The Ramachandran plot analysis by PROCHECK³⁴ showed 73.8%, 20.8%, 4.1% and 1.3% of the protein residues in the most favourable, additionally allowed, generously allowed and disallowed regions, respectively. The ASB

domain alone was crystallized by the hanging-drop vapour diffusion method, against 100 mM Tris, pH 8.5, and 1.5 M LiSO₄. The ASB domain crystals were cryoprotected with 20% glycerol and flash-cooled in a nitrogen stream at 100 K. The native data set was collected at the SPring-8 BL41XU beamline, and the phases were obtained by molecular replacement with MOLREP³⁸. Model building and refinement were performed using Coot³⁹ and Refmac³⁷, respectively.

Kinetic analysis of lysidine modification of pre-tRNA^{Ile2}. In the *G. kaustophilus* genome, the tRNA^{Ile2} gene is located in a cluster of tRNA genes, and is separated from the surrounding tRNA^{Met} and tRNA^{Ser} genes by sequences of 5 and 17 nucleotides, respectively (tRNA^{Met}-TTATC-tRNA^{Ile2}-TTATTATTGTTCTTTT-tRNA^{Ser}). We generated three precursor tRNA^{Ile2} (pre-tRNA^{Ile2}) mutants, by the QuikChange (Stratagene) method. These mutants contained the leader sequence (5L-tRNA^{Ile2}), the trailer sequence (tRNA^{Ile2}-6T) or both the leader and the short trailer sequences (5L-tRNA^{Ile2}-6T). Note that for efficient T7 polymerase *in vitro* transcription, two G residues were added at the beginning of the 5L-tRNA^{Ile2} and 5L-tRNA^{Ile2}-6T mutants. These transcripts and the mature GkRNA^{Ile2} transcript were prepared. GkRNA^{Ile2} lysidinylation assays were performed at 50 °C in 50-μL reaction mixtures, containing 100 mM Tris, pH 8.0, 20 mM KCl, 10 mM MgCl₂, 2 mM ATP, 2 mM DTT, 1 mM spermine, 100 μM ¹⁴C(U)-labelled-L-lysine (Moravek). For the initial velocity measurements, 2 μM of GkRNA^{Ile2} transcripts and 100 nM of GkTlS were used. For the determination of kinetic parameters, 50–500 nM of GkRNA^{Ile2} transcripts and 50–100 nM of GkTlS were used. The filter-binding assay was performed as previously reported²⁷. Kinetic parameters were calculated by fitting to a Hanes-Woolf plot. The assays were performed three times for each tRNA.

***In vitro* lysidine incorporation.** Mutants were generated with a QuikChange kit (Stratagene). The mutant proteins were purified as described for wild-type GkTlS. tRNA^{Met} and tRNA^{Met} from *G. kaustophilus* or *A. aeolicus*, and mutant tRNAs, were purified in the same manner as wild-type GkRNA^{Ile2}. Initial-rate and time-course assays were performed at 60 °C, in the presence of 100 mM Tris-HCl, pH 7.8, 200 mM NaCl, 10 mM MgCl₂, 10 mM KCl, 10 mM DTT, 2 mM ATP, 10.9 mM L-lysine, 175 μM ¹⁴C(U)-labelled-L-lysine (Moravek), 50 μM tRNA transcript and 0.2 μM enzyme. The assay conditions were determined on the basis of the K_m value of the wild type for tRNA.

MD simulation and analyses of the MD trajectory. The details of the methods are provided in the Supplementary Methods.

29. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159 (1987).
30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
31. Weeks, C. M. & Miller, R. The design and implementation of SnB version 2.0. *J. Appl. Cryst.* **32**, 120–124 (1999).
32. de La Fortelle, E., Irwin, J. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
33. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
34. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
36. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
38. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).
39. Emsley, P. C. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

Putting neurons on the map

After a long lull, powerful new technologies are putting the charting of brain circuitry back on neuroscientists' agenda. Michael Eisenstein explores the challenge of mapping the mammalian mind.

It's hard out there for a neuroanatomist — or at least for those who are working to reinvigorate a field that has come to be viewed as outdated and relatively 'unsexy'. "Very few people openly use the term 'neuroanatomy' at this stage for the kind of thing we're talking about, but frankly that's what it is," says Stephen Smith of the Stanford School of Medicine in California. "It's just a kind of neuroanatomy that was impossible until now."

Smith and his colleagues are part of a small community of scientists striving to pick up the mantle of Spanish neuroscience pioneer Santiago Ramón y Cajal by developing sophisticated methods for brain-wide mapping of synaptic connections and neural circuits. "Neuroscience got off to a very good start with the idea that wiring diagrams [are probably the key to understanding] brain function, but remarkably little has happened with that idea," says Jeff Lichtman of Harvard University in Cambridge, Massachusetts.

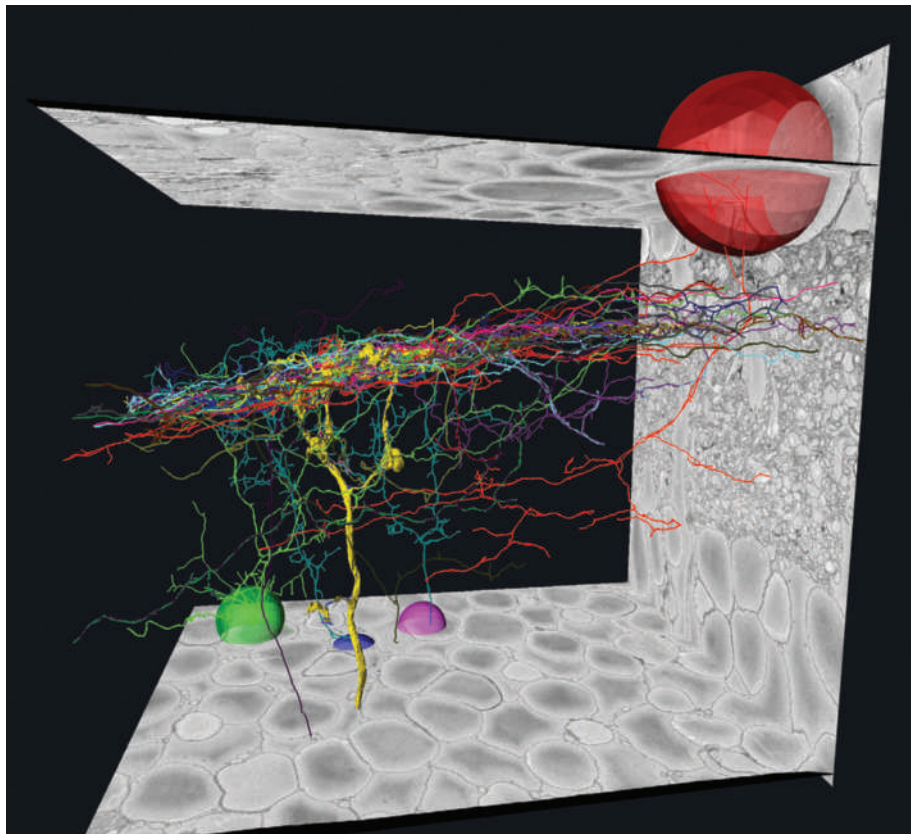
This is now changing, and although these researchers may debate whether to call what they do 'connectomics' or 'circuit mapping' — or even 'neuroanatomy' — there's no question that ongoing strides in cell biology, imaging and computational analysis are bringing scientists closer to understanding the structural foundations of brain function.

Detail-oriented

For more than half a century, scientists have recognized the power that electron microscopy's still-unparalleled resolution could bring to the exploration of neural circuitry. Indeed, the successful assembly by Sydney Brenner and colleagues in the 1980s of the 300-neuron wiring diagram¹ of the nematode worm *Caenorhabditis elegans* was a neuroscience tour de force made possible through reconstruction of transmission electron microscopy (TEM) images from serially collected tissue sections.

Sadly, that was pretty much it for the next 25 years, as the labour-intensive reconstruction process — which consumed more than 10 years' work from Brenner's team — was simply too demanding to deliver large-scale brain mapping. It wasn't until 2004 that Winfried Denk's team at the Max Planck Institute for Medical Research in Heidelberg, Germany, revitalized electron microscopy as a tool for high-resolution neuroanatomy.

In his serial block-face (SBF) imaging method², samples are mounted on an ultramicrotome housed within a scanning electron microscope (SEM), which images the surface of the embedded tissue immediately before



High-resolution data from SBF-SEM enables long-range neuronal reconstructions, such as these cells from the inner-plexiform layer of a rabbit retina.

the diamond knife shaves a thin slice off the top, exposing the next layer for a subsequent round of imaging. This introduces unprecedented capacity for automation to the imaging workflow, but also overcomes several other issues, including the ability to collect data from ultrathin sections without the distortion that can arise with imaging of slices.

Denk's team has since learned how to give the samples a closer shave, boosting the accuracy of reconstruction. "In 2004, the thinnest slices were 40–45 nanometres, but now we're at 25 nanometres," he says. Another major objective has been to tweak staining to optimize the labelling of neuronal processes. "We were working very hard on getting a staining technique that selectively stains the surface of cells and gets rid of the insides, so you don't get distracted by things such as mitochondria, nuclei or the endoplasmic reticulum," says Denk.

This method is compatible with instruments from leading manufacturers such as FEI, of Hillsboro, Oregon, and Hitachi High-Technologies in Tokyo, and users can even buy integrated SBF-SEM systems, such as the 3View

platform made by Gatan in Pleasanton, California, which is based on the Denk lab's design. "I think Denk's was really a landmark publication," says Ben Lich, strategic marketing manager at FEI. At the same time, SBF-SEM is still limited in resolution by the amount of energy that can be pumped into samples safely. "Specimens are typically embedded in a resin, and under the influence of the electron beam they will crosslink and the material becomes harder to cut," explains Lich. "If you put a lot of electrons into your material to create that image, you also do a lot of damage in terms of crosslinking and you cannot really cut it reliably."

As an alternative, FEI is applying technology initially developed for the semiconductor industry, using a focused ion beam to precisely remove thin layers of tissue. "The advantage is that we can put a lot more charge into these blocks and create crosslinking," says Lich, "because the focused ion beam can cut silicon or diamond — basically, we can cut anything with it." Thus, greater resolution is possible, and FEI's DualBeam instruments can image voxels of $4 \times 4 \times 10$ nanometres, relative to the

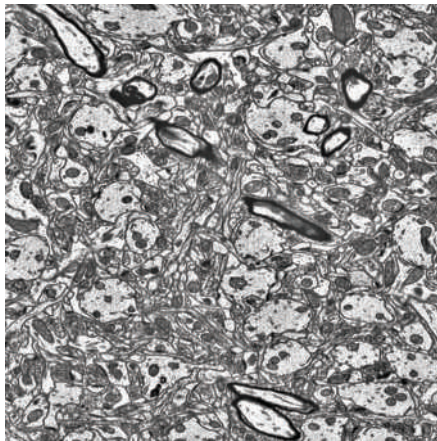
M. HELMSTAEDTER, K. L. BRIGGMAN, & W. DENK

present $20 \times 20 \times 25$ -nanometre resolution limit of SBF-SEM. However, SBF-SEM can image much greater volumes — on the scale of two to three orders of magnitude more — making focused ion beam and SBF complementary rather than competitive tools.

A beautiful mind

There are alternatives to electron microscopy — particularly for researchers interested in more than a static snapshot. “All the things I’ve studied up until now have been dynamic questions,” says Lichtman. “And you just can’t do that with electron microscopy — you’ve got to kill it to look at it!”

Lichtman’s solution was the Brainbow transgenic mouse³, which uses a site-specific DNA recombination system to randomize expression of multiple fluorescent protein genes in neurons, yielding intermediate colour combinations that distinguish each cell from its neighbours. With a broad portfolio of commercially available fluorescent proteins from which to choose — including the Living Colors proteins made by Clontech in Mountain View, California, and the TurboColors proteins from Evrogen in Moscow — Lichtman’s group had many options. However, just a handful of colours proved sufficient to generate nearly 100



Focused-ion-beam microscopy, as performed with instruments such as FEI’s Helios NanoLab DualBeam, allows more energy to be used for imaging, improving the resolution.

distinct labels. “All of the colours of the rainbow that we see are interpreted from three pigments in our retina,” he explains. “So we just inverted that, thinking that if we could just mix different amounts of three colours in different cells, we should be able to get all the visible colours of the rainbow.”

In other cases, more selective labelling is

desirable, and scientists since Ramón y Cajal have pursued chemical and biological methods for exclusively targeting neurons that are functionally linked via active synapses. One promising method, being pioneered by scientists such as Lynn Enquist at Princeton University in New Jersey and Ed Callaway at the Salk Institute for Biological Studies in La Jolla, California, exploits natural infection patterns of neurotropic viruses for the fluorescent labelling of individual neural circuits⁴.

Callaway works with modified rabies virus, a pathogen that spreads so efficiently across mouse neurons that a single particle injected into the brain can prove lethal. His viruses are constrained via deletion of a key glycoprotein gene. “We preserved the ability to replicate and amplify, but provided a means to control the spread,” he says. “Deleting the glycoprotein gene also allows us to control the initial infection and target specific cell types.” Some investigators are applying viral tracing to trace entire networks of interconnected cells, but Callaway is mostly interested in targeting smaller ‘neighbourhoods’. “When we get to the point where we can go into a live animal and target one cell and label every single input to that cell, that will be a huge advance,” says Callaway. “But it’s clear we’re far from labelling all of them. We’re now labelling

C. GENOUD, FMI/FEI

WHOSE MAP IS IT ANYWAY?

Even as ‘connectomics’ makes its way into the mainstream scientific vocabulary, there is already debate over what — if anything — it actually means. “It’s sort of analogous to how ‘genome’ used to mean the set of genes, but now it means the whole DNA sequence,” explains Sebastian Seung of the Massachusetts Institute of Technology in Cambridge.

There is fairly broad agreement that mapping the wiring in mammalian brains is a worthwhile endeavour. The issue is one of scale — should these be comprehensive reconstructions of neuronal circuitry, or more macro-scale representations of long-range connections between regions of the brain? This is the neuroanatomical equivalent of choosing between creating a road atlas or Google Earth.

Arguments can readily be marshalled for and against either approach; most boil down to how best to invest time, money and technology. “Dense reconstruction of a cubic millimetre of the cortex is kind of a ‘going to the Moon’ goal, where we think it’s possible but difficult,” says Seung, “but in

solving that problem, all the other problems become trivial.”

On the other hand, Partha Mitra of Cold Spring Harbor Laboratory in New York thinks that the tools are already at hand for creating a sparser ‘mesoscopic’ map of the projections that link functionally discrete brain regions, which some call a projectome — although you won’t catch Mitra using that term. “Everything has an ‘-ome’ added to it, and that’s ok if you’re in a yoga class,” he jokes. “But I prefer ‘brain architecture’ because it conveys structure and function; architects shape space for human use, and evolution shaped our nervous system for appropriate behavioural repertoires and so on.”

Mitra and dozens of colleagues recently published a plan for integrating existing tools — including chemical labels and engineered viral tracers — into a concerted effort to chart the connections between functionally homogeneous clusters of cells via light microscopy⁹. Larry Swanson at the University of Southern California, Los Angeles, has proposed that 500–1,000 such anatomical units exist within the

brain, and Mitra thinks that linking these will prove challenging but not insurmountable. “Larry has deep knowledge of the relevant literature, and estimates that only around a third of these possible mesoscopic connections have ever been studied,” Mitra says. “But when I sat down and thought about the cost to map out those connections, I was shocked to find that it actually shouldn’t take that much time, money or effort.” With the recent awarding of a Transformative R01 grant from the US National Institutes of Health, Mitra’s team is now taking first steps towards making their Mouse Brain Architecture Project into a reality.

At the same time, by deliberately overlooking the highest orders of neuroanatomical complexity, this approach leaves open numerous questions that will probably be answered only by dense mapping. “These things are just so incredibly tangled and complicated, it’s inconceivable that you’ll come across two identical brains,” says Stephen Smith of Stanford University in California. “And I think it is a wonderful opportunity

for those of us who are pursuing connectomes — to do not one, but many. Neural plasticity is just one of the many interesting questions that will be open for new attack.”

Scientists in both camps hold up work done by researchers at the Allen Institute for Brain Science in Seattle, Washington, in mapping gene expression in the brain as an example of how good science, careful planning and efficient workflows can yield tremendous pay-offs. Smith and others think that maturation of high-resolution circuit-mapping techniques will ultimately bring high-throughput ‘dense’ reconstruction within reach.

Accordingly, Mitra emphasizes that his group’s data — which it intends to make freely available via an open-access model — should provide a framework for future reconstructions. “This is just supposed to be the first generation,” he says. “I have no doubt that if this succeeds — this whole-brain approach to brain architecture and neuroanatomy — then we will see successive waves of technology hitting the problem.”

M.E.

up to 100 inputs, but it should be 1,000.”

By definition, such methods lend themselves to ‘sparse’ mapping of a limited subset of neurons at a lower resolution than ‘dense’ strategies such as electron microscopy. But many scientists see this as a feature rather than a bug, enabling a more selective type of connectomics (see ‘Whose map is it anyway?’) that has the capacity to correlate circuit structure with function. For instance, it can combine the maps with either neuronal activity sensors, such as the calcium-sensitive Cameleon indicator from Invitrogen in Carlsbad, California, or light-activated ion-channel proteins, such as the channel-rhodopsin and halorhodopsin molecules engineered by Karl Deisseroth’s team at Stanford University in California⁵.

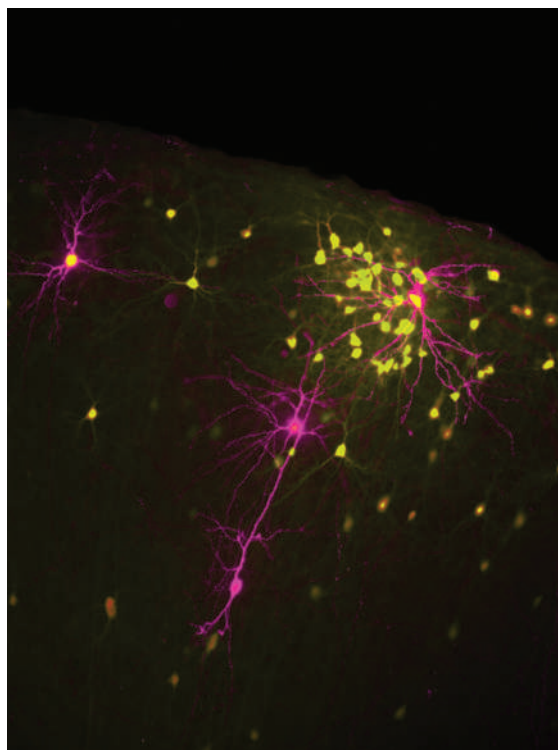
Indeed, even exquisite resolution may soon no longer be the sole domain of electron microscopy, as optical methods emerge that exploit clever workarounds to overcome the diffraction limit for fluorescence imaging, including stimulated emission depletion, stochastic optical reconstruction microscopy, photoactivation localization microscopy and structured illumination⁶, and Lichtman and Smith are among those exploring the benefits and challenges of using ‘super-resolution’ imaging to characterize circuits at the molecular level.

Filling in the details

Another important capability of fluorescence imaging is the ease with which multiple molecular targets can be labelled and readily distinguished — an essential consideration in building a useful map. “As soon as people have a black-and-white connectivity diagram,” says Smith, “they’ll realize they’re really stumped by not knowing what molecules are at a synapse, how the synapse is going to transmit, what its kinetics are going to be, and what’s going to turn it on and off.”

Close pairing of electron microscopy and light microscopy represents a potential solution. In the array tomography technique developed by Smith’s team⁷, for example, a resin-embedded sample is continually sliced by a diamond knife, with the sections sequentially collected on an adhesive surface, enabling them to be arrayed on a slide. These arrays can then be subjected to multiple rounds of immunofluorescence staining and, ultimately, prepared for SEM imaging. Combining imaging modalities enables data relating to expression of channels and receptors to be overlaid onto high-resolution circuit maps. “There are tricks that let us routinely work with 10–15 labels on individual specimens,” says Smith. “It’s a far cry from the 20,000 genes that we’d have to image to fully unlock the brain, but it’s a big step in the right direction.”

Lichtman’s team has been following a similar path, as it continues to refine its automated tape-collecting lathe ultramicrotome (ATLUM)



Ed Callaway’s team is working towards methods to trace every cell that synapses on a target neuron (reprinted with permission from I. R. Wickersham *et al. Neuron* 53, 639–647; 2007).

method⁸. In ATLUM, an epoxy-embedded tissue sample is rotated continuously on a lathe, grazing against a diamond knife that pares away ultrathin sections, which are automatically collected on a continuous strip of adhesive tape. The resulting strips can be imaged by SEM then retained indefinitely for further study. The next generation of this platform promises to deliver sections as thin as 20–25 nanometres along the z-axis, enabling near-seamless circuit

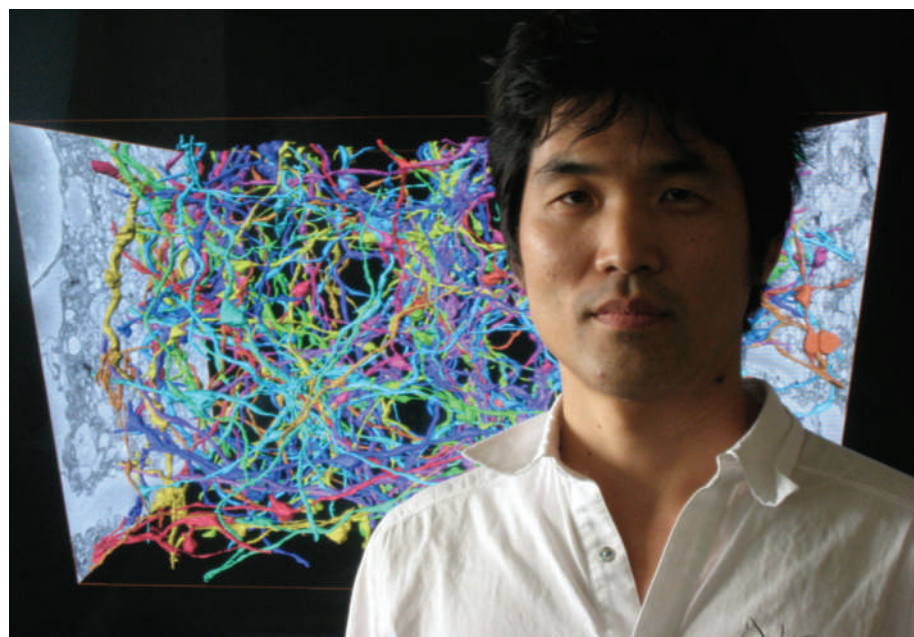
reconstruction. “There is a diminishingly small amount of ambiguity at 30 nanometres, and I think most of those ambiguities would go away if we went down to 25 nanometres,” he says. Lichtman is also looking to integrate fluorescent imaging with ATLUM, perhaps via electron-microscopy-friendly labelling methods that target the tags used in Brainbow. “A purple axon in Brainbow might have two epitope tags, so that blue and red fluorescent proteins can both be stained by immunofluorescence with gold beads of different sizes,” he says. “So if you see an axon with equal numbers of big and little gold beads, then you know that it’s a purple axon.”

This would prove useful not only for integrating data from both modalities, but also as a way to ‘fact-check’ circuit traces. “I think that soon we’ll be able to look at a cell body and predict what kind of glutamate receptors and kinases we’re going to find in the terminals,” says Smith. “So if you follow your wire for millimetres, it better have a certain marker in the synaptic terminal at the end of that wire. If it doesn’t, that means you made a mistake.”

Trace elements

After all the imaging is done, the fundamental problem remains of turning mountains of data into three-dimensional reconstructions and following the neuronal processes that weave through them.

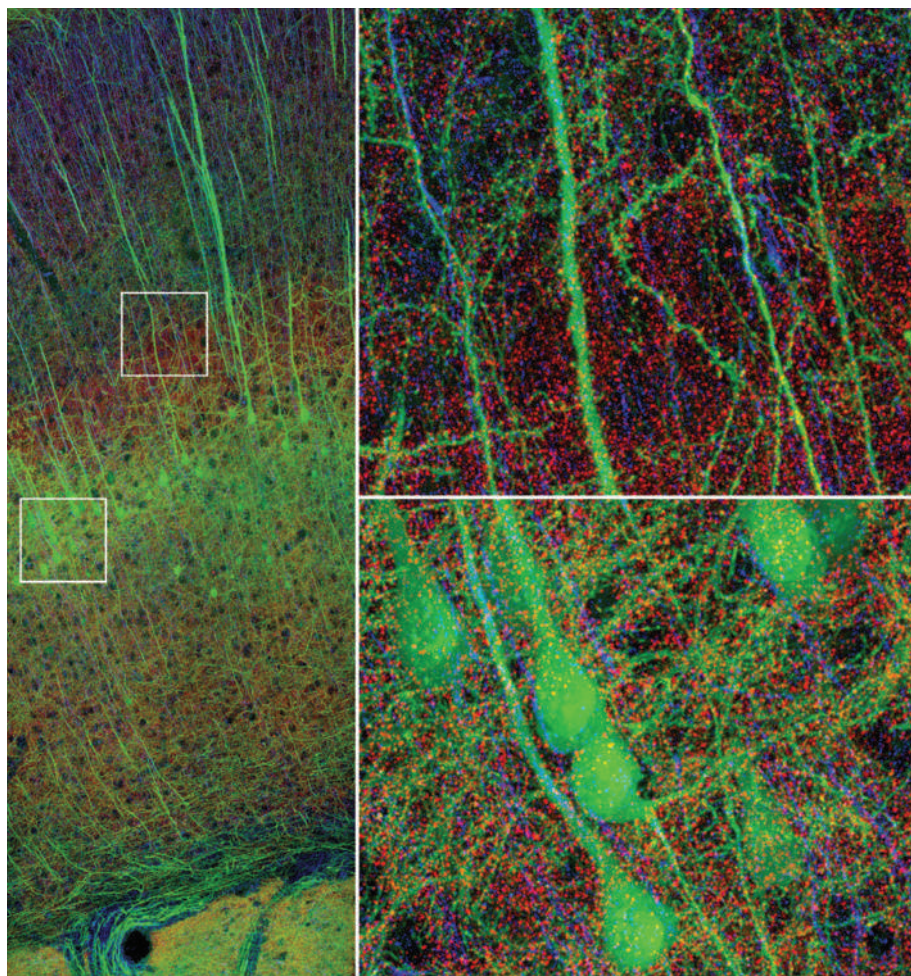
Numerous commercial tools are available for user-guided neuronal tracing, including Imaris from Bitplane in Zurich, Switzerland, and Neurolucida, from MBF Bioscience in Williston, Vermont. Neurolucida was initially developed more than 20 years ago and is one of the most



Sebastian Seung is helping computers to recognize neurons by teaching them to look at images in the same way a human might.

S. SEUNG

S.J. SMITH



Array tomography offers sufficient resolution to distinguish individual synapses, yet also allows direct molecular characterization of cells and connections.

established tools for the manual charting of neurons from fluorescent or electron-microscopy images, offering a relatively straightforward interface for charting complex neuronal processes. “Using a motorized microscope stage and video camera, the software lets you map neuronal processes that are far larger than a single field of view,” explains chief scientific applications officer Geoff Greene. “When the user reaches the end of a dendrite, the software will automatically take you back to the x - y - z coordinates of the last unfinished branch, so it can travel down the alternate branch and trace it.” Denk’s team has also developed an elegant tool, Knossos, for rapidly tracing neurons within their reconstructions, in which users sketch rudimentary ‘skeletons’ along the path taken by a given neurite.

Of course, manual tracing is wholly impractical for large-scale mapping, and the hunt is on for algorithms that can automatically define individual neurons within three-dimensional reconstructions, a process known as segmentation. Neurolucida features a module called AutoNeuron that strives to deliver rapid, machine-assisted tracing, but Greene acknowledges that this is still a work in progress. The fundamental problem is that near-perfect accuracy is required. “If you lose a wire some-

where along its length, then you lose all the connections downstream from where you lose it,” says Denk. “If I lose an axon halfway down its length, I could lose 5,000 connections.”

The core of the segmentation problem, explains Sebastian Seung of the Massachusetts Institute of Technology in Cambridge, is getting computers to see the world as humans do. “To be able to trace the neurons, we have to know the boundaries of every object,” he says. “This is one of the first problems ever attacked in computer vision, back in the late 1960s — and yet today we still don’t have reliable systems that do it.”

Rather than trying to establish strict neuron-recognition guidelines, Seung and his col-



Partha Mitra thinks that scientists already have the tools they need to begin building a ‘mesoscale’ map of the mouse brain.

leagues such as Dmitri Chklovskii at the Janelia Farm Research Campus in Ashburn, Virginia, are pursuing ‘machine-learning’ strategies that teach computers by example. “We have humans trace the boundaries, and create a training set,” says Seung, “then we have the computer learn how to imitate the human tracing.” This approach brings with it a number of complex challenges, including the development of metrics that enable unambiguous comparison of different tracing strategies and quantification of overall accuracy. These computer ‘students’ also need to know the limitations of their organic instructors. “When humans trace stuff, there’s just a lot of jitter in the way they trace,” says Seung. “So we devised ways in which computers can be made to imitate humans but not take them too literally.”

Even historically pure ‘wet labs’ are finding themselves joining the fight. Lichtman’s lab routinely churns out hundreds of $16,000 \times 16,000$ -pixel images, each of which can consume a gigabyte of space; as such, developing tools for data handling is now a daily fact of life. “In my hiring now, about a third of the people I’m looking at are people who are computer scientists,” he says, “and that would have been unthinkable five years ago.”

Thinking big

Ultimately, many of the obstacles that constrain large-scale circuit mapping boil down to maximizing throughput; for example, the rate of electron-microscopy imaging. “We’ll either have to parallelize acquisition, or have a microscope that’s much faster,” says Denk. “And I don’t mean by a factor of five, I mean by factors of 100.”

Consistent quality of sample preparation is also a key problem, as all downstream analysis rests on this step. Accordingly, Smith’s main nemesis these days is the dirt that can obliterate fine details from specimens, and his team is learning lessons in cleanliness from their neighbours in Silicon Valley. “I take much heart from the fact that today people can make a microprocessor chip with a billion transistors that each work perfectly for about \$20 a chip,” he says.

In fact, many cite the semiconductor industry as a model for what will be needed for any large-scale ‘Connectome Project’: consistent application of an established set of optimized methods. “At some point we’ll have all the tools lined up,” says Denk, “and then we’ll decide to spend some real money on this to do the whole brain of some animal.”

Michael Eisenstein is a writer based in Brooklyn, New York.

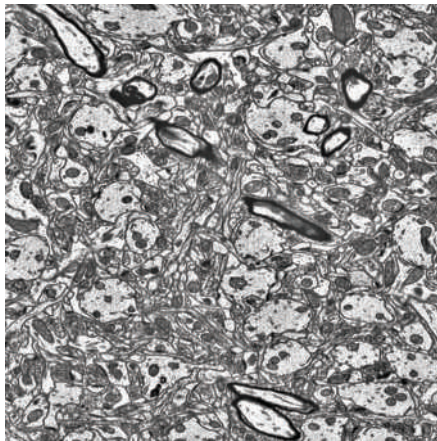
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Denk, W. & Horstmann, H. *PLoS Biol.* **2**, e329 (2004).
- Livet, J. et al. *Nature* **450**, 56–62 (2007).
- Callaway, E. M. *Curr. Opin. Neurobiol.* **18**, 617–623 (2008).
- Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L. & Deisseroth, K. *Nature Rev. Neurosci.* **8**, 577–581 (2007).
- Rae Chi, K. *Nature Meth.* **6**, 15–18 (2009).
- Micheva, K. D. & Smith, S. J. *Neuron* **55**, 25–36 (2007).
- Hayworth, K. J., Kasthuri, N., Schalek, R. & Lichtman, J. W. *Microsc. Microanal.* **12**, Suppl. S02, 86–87 (2006).
- Bohland, J. W. et al. *PLoS Comput. Biol.* **5**, e1000334 (2009).

present $20 \times 20 \times 25$ -nanometre resolution limit of SBF-SEM. However, SBF-SEM can image much greater volumes — on the scale of two to three orders of magnitude more — making focused ion beam and SBF complementary rather than competitive tools.

A beautiful mind

There are alternatives to electron microscopy — particularly for researchers interested in more than a static snapshot. “All the things I’ve studied up until now have been dynamic questions,” says Lichtman. “And you just can’t do that with electron microscopy — you’ve got to kill it to look at it!”

Lichtman’s solution was the Brainbow transgenic mouse³, which uses a site-specific DNA recombination system to randomize expression of multiple fluorescent protein genes in neurons, yielding intermediate colour combinations that distinguish each cell from its neighbours. With a broad portfolio of commercially available fluorescent proteins from which to choose — including the Living Colors proteins made by Clontech in Mountain View, California, and the TurboColors proteins from Evrogen in Moscow — Lichtman’s group had many options. However, just a handful of colours proved sufficient to generate nearly 100



Focused-ion-beam microscopy, as performed with instruments such as FEI’s Helios NanoLab DualBeam, allows more energy to be used for imaging, improving the resolution.

distinct labels. “All of the colours of the rainbow that we see are interpreted from three pigments in our retina,” he explains. “So we just inverted that, thinking that if we could just mix different amounts of three colours in different cells, we should be able to get all the visible colours of the rainbow.”

In other cases, more selective labelling is

desirable, and scientists since Ramón y Cajal have pursued chemical and biological methods for exclusively targeting neurons that are functionally linked via active synapses. One promising method, being pioneered by scientists such as Lynn Enquist at Princeton University in New Jersey and Ed Callaway at the Salk Institute for Biological Studies in La Jolla, California, exploits natural infection patterns of neurotropic viruses for the fluorescent labelling of individual neural circuits⁴.

Callaway works with modified rabies virus, a pathogen that spreads so efficiently across mouse neurons that a single particle injected into the brain can prove lethal. His viruses are constrained via deletion of a key glycoprotein gene. “We preserved the ability to replicate and amplify, but provided a means to control the spread,” he says. “Deleting the glycoprotein gene also allows us to control the initial infection and target specific cell types.” Some investigators are applying viral tracing to trace entire networks of interconnected cells, but Callaway is mostly interested in targeting smaller ‘neighbourhoods’. “When we get to the point where we can go into a live animal and target one cell and label every single input to that cell, that will be a huge advance,” says Callaway. “But it’s clear we’re far from labelling all of them. We’re now labelling

C. GENOUD, FMI/FEI

WHOSE MAP IS IT ANYWAY?

Even as ‘connectomics’ makes its way into the mainstream scientific vocabulary, there is already debate over what — if anything — it actually means. “It’s sort of analogous to how ‘genome’ used to mean the set of genes, but now it means the whole DNA sequence,” explains Sebastian Seung of the Massachusetts Institute of Technology in Cambridge.

There is fairly broad agreement that mapping the wiring in mammalian brains is a worthwhile endeavour. The issue is one of scale — should these be comprehensive reconstructions of neuronal circuitry, or more macro-scale representations of long-range connections between regions of the brain? This is the neuroanatomical equivalent of choosing between creating a road atlas or Google Earth.

Arguments can readily be marshalled for and against either approach; most boil down to how best to invest time, money and technology. “Dense reconstruction of a cubic millimetre of the cortex is kind of a ‘going to the Moon’ goal, where we think it’s possible but difficult,” says Seung, “but in

solving that problem, all the other problems become trivial.”

On the other hand, Partha Mitra of Cold Spring Harbor Laboratory in New York thinks that the tools are already at hand for creating a sparser ‘mesoscopic’ map of the projections that link functionally discrete brain regions, which some call a projectome — although you won’t catch Mitra using that term. “Everything has an ‘-ome’ added to it, and that’s ok if you’re in a yoga class,” he jokes. “But I prefer ‘brain architecture’ because it conveys structure and function; architects shape space for human use, and evolution shaped our nervous system for appropriate behavioural repertoires and so on.”

Mitra and dozens of colleagues recently published a plan for integrating existing tools — including chemical labels and engineered viral tracers — into a concerted effort to chart the connections between functionally homogeneous clusters of cells via light microscopy⁹. Larry Swanson at the University of Southern California, Los Angeles, has proposed that 500–1,000 such anatomical units exist within the

brain, and Mitra thinks that linking these will prove challenging but not insurmountable. “Larry has deep knowledge of the relevant literature, and estimates that only around a third of these possible mesoscopic connections have ever been studied,” Mitra says. “But when I sat down and thought about the cost to map out those connections, I was shocked to find that it actually shouldn’t take that much time, money or effort.” With the recent awarding of a Transformative R01 grant from the US National Institutes of Health, Mitra’s team is now taking first steps towards making their Mouse Brain Architecture Project into a reality.

At the same time, by deliberately overlooking the highest orders of neuroanatomical complexity, this approach leaves open numerous questions that will probably be answered only by dense mapping. “These things are just so incredibly tangled and complicated, it’s inconceivable that you’ll come across two identical brains,” says Stephen Smith of Stanford University in California. “And I think it is a wonderful opportunity

for those of us who are pursuing connectomes — to do not one, but many. Neural plasticity is just one of the many interesting questions that will be open for new attack.”

Scientists in both camps hold up work done by researchers at the Allen Institute for Brain Science in Seattle, Washington, in mapping gene expression in the brain as an example of how good science, careful planning and efficient workflows can yield tremendous pay-offs. Smith and others think that maturation of high-resolution circuit-mapping techniques will ultimately bring high-throughput ‘dense’ reconstruction within reach.

Accordingly, Mitra emphasizes that his group’s data — which it intends to make freely available via an open-access model — should provide a framework for future reconstructions. “This is just supposed to be the first generation,” he says. “I have no doubt that if this succeeds — this whole-brain approach to brain architecture and neuroanatomy — then we will see successive waves of technology hitting the problem.”

M.E.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Microscopy systems			
Applied Precision	Fluorescence microscopes and the structured-illumination-based OMX system	Issaquah, Washington	www.api.com
ASPEX	SEM instruments and software	Delmont, Pennsylvania	www.aspexcorp.com
Carl Zeiss	Confocal, multiphoton and electron microscopes, and accessories	Oberkochen, Germany	www.zeiss.com
Delong Instruments	Electron microscopes, guns and detectors	Brno, Czech Republic	www.dicomps.com
FEI	TEM, SEM and focused ion beam (FIB) instruments	Hillsboro, Oregon	www.fei.com
FOCUS	Photoemission electron microscopy (PEEM) instruments	Hünstetten-Kesselbach, Germany	www.focus-gmbh.com
Gatan	Instruments and software for TEM, SEM and SBF-SEM	Pleasanton, California	www.gatan.com
Hitachi High Technologies	TEM, SEM and FIB instruments	Tokyo, Japan	www.hitachi-hitec.com
Intelligent Imaging Innovations	Confocal and multiphoton microscopes and software	Denver, Colorado	www.intelligent-imaging.com
ISS Group	New and pre-owned electron-microscopy instruments and accessories	Withington, UK	www.iss-group.co.uk
JEOL	Instruments and tools for TEM, SEM and FIB	Tokyo, Japan	www.jeol.com
Kingstic Optical	Light microscopes and accessories	Ningbo, China	www.biomicroscope.com
Leica Microsystems	Confocal microscopes, ultramicrotomes for SEM sample preparation	Wetzlar, Germany	www.leica-microsystems.com
Micro Video Instruments	Distributor of light microscopes and imaging hardware from various manufacturers	Avon, Massachusetts	www.mvi-inc.com
Nikon Instruments	Confocal microscopes, SEM and software	Melville, New York	www.nikoninstruments.com
Olympus	Confocal, multiphoton, SEM and TEM systems	Tokyo, Japan	www.olympus-global.com
SEMtech Solutions	Second-hand SEM instruments and accessories	North Billerica, Massachusetts	www.semtechsolutions.com
Tescan	New and used SEM instruments	Brno, Czech Republic	www.tescan.com
Visitec	Large-chamber SEM and accessories	Grevesmühlen, Germany	www.visitec-em.de
Electron-microscopy sample preparation and accessories			
Campden Instruments	Vibrating microtomes and blades	Loughborough, UK	www.campden-inst.com
Cressington	Sample-coating systems for SEM and TEM	Watford, UK	www.cressington.com
Diatome	Diamond knives	Biel, Switzerland	www.diatome.ch
Electron Microscopy Sciences	Vibrating microtomes and other tools for SEM and TEM sample preparation	Hatfield, Pennsylvania	www.emsdiasum.com
Energy Beam Sciences	Sample preparation and accessories for SEM	East Granby, Connecticut	www.ebsciences.com
Fischione Instruments	Hardware for EM sample preparation	Export, Pennsylvania	www.fischione.com
KE Developments	Detectors, chamberscopes, and other SEM accessories	Cambridge, UK	www.kedev.co.uk
M.E. Taylor Engineering	Scintillators and sample mounts for SEM	Brookeville, Maryland	www.semico.org
Micro Star Technologies	Diamond knives	Huntsville, Texas	www.microstartech.com
Precisionary Instruments	Compressotome vibrating microtome	Greenville, North Carolina	www.precisionary.com
Quorum Technologies	Carbon and sputter coaters and other sample preparation	East Grinstead, UK	www.quorumtech.com
RMC Products	Microtomes and ultramicrotomes	Tucson, Arizona	www.rmcprouducts.com
South Bay Technology	Sample preparation for SEM and TEM	San Clemente, California	www.southbaytech.com
SPI Supplies	Chemicals and stains for EM	West Chester, Pennsylvania	www.2spi.com
Ted Pella	Hardware and supplies for sample preparation	Redding, California	www.tedpella.com
Thermo Scientific	Microtomes	Waltham, Massachusetts	www.thermo.com
Tousimis	Critical point dryers, sample holders and other accessories	Rockville, Maryland	www.tousimis.com
Vibratome	Vibratomes and other microtomes	St Louis, Missouri	www.vibratome.com
Optics and cameras			
Andor Technology	CCD, EMCCD and ICCD cameras	Belfast, Northern Ireland	www.andor.com
Chroma Technology	Optical filters	Rockingham, Vermont	www.chroma.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Cooke	Diverse cameras for fluorescence imaging	Romulus, Michigan	www.cookecorp.com
Diagnostic Instruments	SPOT family of cameras	Sterling Heights, Michigan	www.diaginc.com
DVC	Digital cameras and software	Austin, Texas	www.dvcco.com
Graftek Imaging	Digital cameras and software	Austin, Texas	www.graftek.com
Hamamatsu	Cameras and detectors	Hamamatsu City, Japan	www.hamamatsu.com
Omega Optical	Optical filters	Brattleboro, Vermont	www.omegafilters.com
Optronics	Microscope cameras and imaging software	Goleta, California	www.optronics.com
Princeton Instruments	Cameras and laser optics	Trenton, New Jersey	www.princetoninstruments.com
QImaging	Various CCD cameras	Surrey, British Columbia	www.qimaging.com
Semrock	Optical filters and mirrors	Rochester, New York	www.semrock.com

Imaging software

4pi Analysis	Digital imaging tools for SEM and TEM	Durham, North Carolina	www.4pi.com
Amerinex Applied Imaging	Aphelion image processing software	Monroe, New Jersey	www.amerineximaging.com
Bitplane	Imaris imaging software suite	Zurich, Switzerland	www.bitplane.com
Improvision	Volocity three-dimensional reconstruction software	Coventry, UK	www.improvision.com
MathWorks	MATLAB programming environment	Natick, Massachusetts	www.mathworks.com
MBF Bioscience	Neurolucida imaging software suite	Williston, Vermont	www.mbfbioscience.com
MDS Analytical Technologies	Metamorph image-processing software	Concord, Ontario	www.moleculardevices.com
MediaCybernetics	Image Pro Plus processing software	Bethesda, Maryland	www.mediacy.com
Scientific Volume Imaging	Huygens deconvolution software	Hilversum, The Netherlands	www.svi.nl
Visage Imaging	Amira software for image reconstruction	Richmond, Australia	www.visageimaging.com
Visiopharm	Visiopharm Integrator System software	Hørsholm, Denmark	www.visiopharm.com

Labels, stains and dyes

AnaSpec	Antibodies, fluorescent dyes and labelling kits	Fremont, California	www.anaspec.com
Biotium	Dyes, fluorescent indicators and antibodies	Hayward, California	www.biotium.com
Clontech	Fluorescent proteins and reporters	Mountain View, California	www.clontech.com
Denovo Biolabels	Cyanine-based fluorophores	Münster, Germany	www.biolabels.com
eBioscience	Organic dyes and fluorescent nanocrystals	San Diego, California	www.ebioscience.com
EMD Chemicals	Dyes, stains and antibodies	Gibbstown, New Jersey	eshop.emdchemicals.com
Evrogen	Fluorescent proteins and reporters	Moscow, Russia	www.evrogen.com
Invitrogen	Chemical and genetically encoded fluorescent labels and indicators	Carlsbad, California	www.invitrogen.com
MBL International	Conventional and photoswitchable fluorescent proteins	Woburn, Massachusetts	www.mblintl.com
MoBiTec	Fluorescent dyes and indicators	Göttingen, Germany	www.mobitec.de
Polysciences	Chemical labelling and staining reagents	Warrington, Pennsylvania	www.polysciences.com
Sigma-Aldrich	Antibodies and labelling reagents	St Louis, Missouri	www.sigmaaldrich.com
Tocris Bioscience	Fluorescent probes and activity indicators	Ellisville, Missouri	www.tocris.com

● see advertisement

PROSPECTS

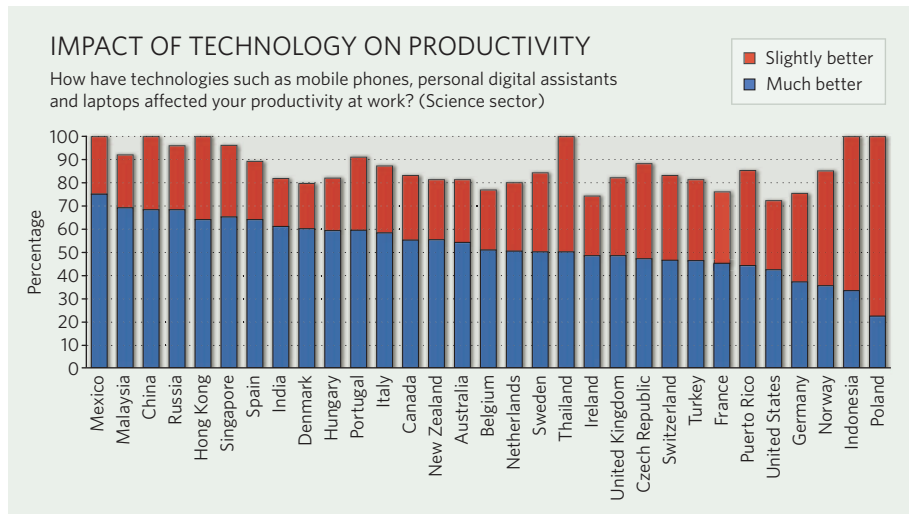
Going mobile

Survey results suggest that mobile technology offers scientists both increased productivity and unwelcome intrusion. Rich Pennock speculates on the consequences.

Mobile communications have transformed the average workplace, and the science workplace is no exception. The potential to work remotely has led to a higher level of job satisfaction and a sense of increased productivity, according to the Kelly Global Workforce Index, a survey conducted by workforce-management company Kelly Services. But this boost in connectivity and productivity has taken a serious toll on scientists' work-life balance.

Nearly 100,000 professionals, including more than 3,000 in the international science community, responded to the survey across North America, Europe and the Asia Pacific region. Questions focused on respondents' work-life balance and ways in which technology has affected their lives and careers.

The positive results: a big boost in perceived productivity. More than 70% of science professionals surveyed said that the ability to remain in constant communication with their work is a positive development. And 80% believe that they are more productive now than they were before the advent of mobile communications technology. Yet this sense of higher productivity comes at a price, the survey found. More than a third of science professionals polled worldwide are dissatisfied with their current work-life balance.



Telecommuting has hastened the disappearance of the line separating work and home. Eighty-three per cent of the 3,000 scientist respondents said that a workplace policy on telecommuting is an attractive component in deciding where to work. But with laptops and smartphones always within arm's reach, almost 30% of the scientists surveyed said that they are working more hours than they used to.

Employers must forge a balance. According to survey findings, employees who have

a good work-life balance tend to be more productive, more engaged and happier in their work. To maintain employee morale and productivity, managers should provide rewards to employees for good performance — for example, extra holiday time, or restaurant or entertainment coupons. Happy researchers are those who not only have access to the technology, but also the means and the incentives to completely disconnect. ■

Rich Pennock is vice-president of Kelly Services in Troy, Michigan.

Q&A

Stefan Söldner-Rembold, a particle physicist at the University of Manchester, UK, is the latest spokesperson elected to co-ordinate the D0 experiment, an exploration of the subatomic universe that started in 1992 at the Fermi National Accelerator Laboratory in Batavia, Illinois.



As the second non-US scientist to hold the spokesperson post, what challenges do you face?

I've been at D0 since 2001. Since then, non-US contribution to D0 has increased significantly. Currently, about half the physicists on the experiment are from outside the United States. The challenge is navigating the different academic cultures and funding schemes to bring these groups together.

What has been your most exciting scientific endeavour?

At the former Large Electron-Positron collider at CERN [the particle-physics lab near Geneva, Switzerland], and now at the

Tevatron accelerator at Fermilab, our research has focused on proving or disproving the existence of the Higgs boson, an elusive particle whose predicted existence may explain how elementary particles acquire mass. It is the most exciting and important thing I've been involved with during my career.

D0 is set to end in 2010; will you be the last spokesperson appointed?

Not necessarily. We are approved until 2010 and expect to run through most of 2011. Currently, we have taken only half the data we expect to take by 2011. The machine has improved so

much over the years, resulting in many more proton-antiproton collisions and leading us to take ten times more data now than at first. So data analysis from D0 will continue for many years.

So D0 is still quite active?

Yes. In many areas of research at D0, we've published only 20% of what we expect to. There is a lot of physics out there — which is exciting for us and our students.

Has the Large Hadron Collider (LHC) usurped talent from D0?

The main challenge is balancing manpower between the Tevatron, currently the highest-energy running particle accelerator in

the world, and the start-up of the LHC experiments [at CERN], as many of us work on both LHC and Tevatron experiments. However, any competition is positive. Success of the LHC is extremely important for the field of particle physics. We are all in the same boat.

What is the main goal for D0 before it ends?

As theoretical predictions point to a Higgs boson in a mass range accessible to the Tevatron, we may be able to exclude its existence in this mass range or see the first evidence for its existence. ■

Interview by Virginia Gewin

Cloud Gate sculpture
in downtown Chicago.

Seeds of collaboration

Academic and government labs in the Chicago area are combining forces to reel in a host of large collaborative research projects — and tens of millions of dollars in funding. **Paul Smaglik** sums up.

Chicago, popularly known for its gangsters, sports fans and deep-dish pizza, also has a cluster of impressive research institutions that often get overshadowed by the elite universities on both US coasts. But now, the city's biggest research players are beginning to team up, combining their time, talent and collective strengths to earn serious federal funding.

In the past year, the Argonne National Laboratory, the University of Illinois at Chicago (UIC), the University of Chicago and Northwestern University have landed four of 46 Department of Energy (DOE) 'Energy Frontier Research Center' (EFRC) grants worth \$19 million apiece, as well as several US National Institutes of Health (NIH) and National Science Foundation (NSF) awards worth \$10 million or more. These awards alone should create a few hundred jobs in Chicago, thus cementing the city's recent successes in garnering science funding across multiple disciplines.

The state of Illinois, and Chicago in particular, has succeeded by diversifying its research portfolio — in 2008, it ranked ninth in landing NIH money in the top 50 institutions for grants nationwide (UIC ranked 48th with \$134 million, Northwestern came in 37th with \$168 million and the University of Chicago was placed 28th with \$194 million). Illinois also ranked fifth in

terms of both NSF (\$257 million) and DOE grants (\$30.8 million).

These institutions have historically done well in securing federal research funding. But recent successes in landing large collaborative grants that have boosted each university's total have come about very much by design. In 2002, a philanthropic fund set up by the Searle family, which built its fortune on pharmaceuticals such as the first oral contraceptive and the anti-inflammatory drug Celebrex, saw an opportunity to turn rival institutions into collaborators. The family formed the Chicago Biomedical Consortium (CBC), supplied it with \$5 million and mandated funding partnerships.

The CBC has dedicated the money to symposia, supplemental grant money and shared facilities. These facilities include a proteomics centre at UIC, a cellular screening centre at the University of Chicago and a mass spectrometry centre at Northwestern. Yearly CBC-sponsored symposia allow researchers from each of the universities to meet and plan for joint projects. And CBC funds help support collaborative research by topping up successful grants that span multiple Chicago research institutions. The CBC has caused a cultural shift in Chicago's research

landscape, says CBC scientific director Jonathan Silverstein. "For a long time, these were essentially competing institutions," Silverstein says.

The collaborations help because many projects simply cannot be done in isolation. "There are problems that are sufficiently large and sufficiently complex that one person, one department, one institution can't do them," says Donald Levy, vice-president of research at the University of Chicago. Chicago already has some glue that can help hold such large projects together, he says — Argonne. The national laboratory possesses both the

supercomputing power, such as Blue Gene, and expensive scientific equipment, such as the Advanced Light Source synchrotron, that can serve as common ground for structural biology and materials science research.

In August, that glue got stronger. The NSF awarded

\$30 million to improve TeraGrid, a distributed network of 11 supercomputing sites. One of the grant's two recipients, Ian Foster, directs the Computation Institute, which is split between Argonne and the University of Chicago. The grant will allow scientists at each of the sites to increase the network's speed and usability.

Users can book computing time, then

"None of our departments are big enough that they can stand alone."
— Jay Walsh

J. FUSTER RAGA/CORBIS

tap into the grid's power to perform complex calculations and visualizations, from simulating earthquakes to modelling human blood flow and the impact of plaque formation in the arterial system. "Data-intensive science is an area where Chicago seems to have a critical mass of people," Foster says. Having ready access to supercomputing nodes acts like a magnet to make that mass denser and stronger.

For Kevin White, having accessible computing power was a necessity. The first professor to receive CBC funding, White, the director of the Institute of Genomics and Systems Biology at the University of Chicago, received a joint appointment with Argonne, which would provide him the resources necessary to do his computation-heavy work. They gave him \$1 million to start his lab. Then, when he landed a \$9.1-million grant from the NIH's Human Genome Research Institute to identify the transcription factors that control how and when genes express proteins, the CBC gave him a 'lever grant' of \$3 million to build research tools that will benefit the other universities. That leverage paid off last year with a five-year, \$15-million grant from the NIH's National Institute of General Medical Sciences to become one of 10 NIH-designated systems biology institutes.

The CBC provided him with plenty of space and the NIH with plenty of money. Now White looks to fill rows of empty benches and computer terminals — some with views of Lake Michigan — with a mix of computational and experimental biologists. White expects to grow the number of principal investigators from four now to 20 in a few years. Meanwhile, he's working with other groups both in the United States and abroad. With the European Molecular Biology Laboratory in Heidelberg, Germany, he helped develop a technology where researchers can make 'movies' of real-time gene expression in organisms under microscopes. Closer to home, he's working with the University of Chicago Cancer Center, Northwestern University and UIC to create expression profiles of breast-cancer tumours.

Northwestern, too, has benefited from working with its former competitors, says Jay Walsh, Northwestern's vice-president for research. The university's need to find research collaborators was born of necessity, with much of its science faculty 'crammed' into a relatively small corner of campus. "None of our departments are big enough that they can stand alone," Walsh says, "When faculty look to collaborate, they have to look outside."

The university had another opportunity to form an external partnership when, four years ago, Northwestern, the University of Chicago and UIC began jointly managing Argonne. "What that does is it brings us

"You look upon the other guy's success as your success."
— Donald Levy



ARGONNE NAT LAB

Illinois's Argonne National Laboratory has Blue Gene (above) supercomputing power at its fingertips.

together on a regular basis and binds us," says Walsh. It also allows for individual expertise to be both shared and polished. For example, Wayne Anderson has secured the university's (NSF's) single biggest grant — \$30 million to solve protein structures, using high-throughput technology to crystallize them, Argonne's Advanced Photon Source synchrotron to probe them and the

Computation Institute to model their three-dimensional shapes.

Northwestern's relationship with Argonne has helped the university win two EFRC awards, each worth \$19 million. Both are seeking cheaper, cleaner ways to create energy.

Each of the centres will probably result in the hiring of 25–30 new personnel, according to Michael Wasielewski, a Northwestern chemistry professor who landed one of the EFRC grants. Wasielewski already works with Argonne scientists on the Argonne-Northwestern Solar Energy Research (ANSER) Center. The grant will enable his team to analyse photosynthesis for ways to create more efficient photovoltaic cells and create hybrid solar cells that have both organic and inorganic components. "The people involved in this now are really serious about solving the problem," he says. "They're not looking at this as just another funding source."

Ironically, Wasielewski has some friendly competition. His Northwestern colleague Bartosz Grzybowski, associate professor of chemical and biological engineering, received a similar EFRC grant — and is taking a completely different approach. Wasielewski looks to improve photovoltaic cells in small increments rather than in seemingly impossible leaps. Grzybowski seeks something more dramatic — harnessing the power of materials under conditions far from equilibrium. "It might be in the category of crazy," Grzybowski says.

"But I don't think it is."

Crazy or not, Grzybowski's approach requires recruiting lots of brains. "The biggest chunk of money is going to people," he says. Grzybowski aims to recruit the top chemists in the world to his effort — especially ones with nanotechnology experience. Mary Jo LaDu, associate professor of anatomy and cell biology in the UIC College of Medicine, is also looking beyond the boundaries of Chicago to staff her \$10-million project to characterize the apoE protein, thought to play a central role in Alzheimer's disease. LaDu received the \$10 million grant on 15 August and has already assembled a dream team of researchers from UIC, the University of Kentucky, Washington University School of Medicine in St Louis, Georgetown University in Washington and the University of South Florida in Tampa. A transgenic mouse she developed serves as a binding component in the collaboration.

But LaDu is going beyond her mouse model to create a more cohesive apoE community. The collaboration will include a yearly symposium at varying sites that will give young, emerging researchers almost as much time to talk as keynote speakers. LaDu is also building a website that will allow researchers outside the grant to access the group's reagents, techniques and findings. Her goal for the website? To ask participants "What do you know? What do you have? What do you think?" says LaDu.

All the new collaborative grants should collectively bring hundreds of new researchers into the area. And the team approach — fostered by the CBC — is changing the way these groups interact with each other. "The old philosophy was 'Oh my God! The other guy is beating us,'" says Levy. The new one turns former competitors into cheerleaders. "You look upon the other guy's success as your success," Levy says.

Paul Smaglik is a freelance writer in Milwaukee, Wisconsin.

Tropicbird

The winds of change.

KV

I don't remember whose idea it was to follow the tropicbird. Perhaps it was mine — a vain hope to chase the memories of my coastal childhood under the pretence of observing an odd visitor to the bay. Perhaps it was Hector's idea: Hector, who was afraid of nothing that the increasingly violent and unpredictable Gulf could produce, in the early years of that storm-wracked decade. Hector, whose catamaran, the *Danaus*, was rarely tethered in its slip at the university harbour for more than two days at a stretch. In the iterated dream, he turns his handsome face to smile at me, and releases the sail of the *Danaus* as the tropicbird burns white against the bruised horizon.

The destruction of Gulf Coast communities, a consequence of anthropogenic climate change, began with Hurricane Katrina in 2005 and Hurricane Ike in 2008. By 2015, even the sceptics could not deny the dramatic increases in the annual number and strength of hurricanes in the Gulf of Mexico. Despite extensive improvements achieved during President Obama's third term, the New Orleans levees failed in 2023, and the resulting flood made the city uninhabitable. Five years later, only the tops of the high-rise buildings were visible above the waves, and Lake Pontchartrain, along with the crescent bend of the Mississippi River, had been swallowed by the Gulf. Galveston, island site of a bio-terrorism research facility, was destroyed by Hurricane Omicron in 2032, the year I was born. By that time, the hurricanes inevitably killed so many people each year, that no one had the heart to give them familiar human names.

We met as undergraduates in Houston, both of us seeking degrees in environmental science. Hector, consumed with youthful wanderlust, was interested in how climate change and expansion of the Gulf affected the patterns of animal migrations, particularly for the monarch butterfly. One had to be energetic to engage in ecological research during those oil-impooverished years. Our older professors regaled us with tales of gas-powered vehicles, used routinely in field work during the early decades of the twenty-first century, and our eyes widened in disbelief. We knew only the immediate world, in which remote research stations were reached on horseback, and coastal sites accessed by kayak

or sailboat. Our youthful restlessness and romantic notions of science drove us out into the field, to conduct studies of a coastal environment that changed rapidly with the climate. It would not be an exaggeration to label our university years as halcyon ones for environmental research: funding was available from federal agencies, and a number of solar-powered instruments allowed us to continue imaging and molecular investigations in the laboratory. On that October day, with our senior projects focused on climate influences on seabirds and butterflies, Hector and I were eager to



begin mapping the impact of a Category 1 hurricane on our research populations. It was predicted to make landfall east of Houston, and its path lay within reach of our boats.

No traces of the *Danaus*, or of Hector, were ever discovered by the Coast Guard. Hurricane Laguz blew ashore east of Oak Island and Double Bayou, dissipating as it swirled northward, over the abandoned interstate highway. I tried to explain to the authorities who questioned me that Hector was an experienced sailor, familiar with the mood swings and sudden squalls of the Gulf, aware of the submerged buildings. He had been monitoring the migratory streams of monarchs for several years, and routinely collected data about wind speed and direction, currents, temperatures, tides and man-made obstacles. The hurricane coincided with the southward autumn journey of the monarchs, to the great trees in the mountains of Michoacán. I had followed Hector's catamaran in my kayak until the waves had threatened to

swamp the craft. I mentioned neither the tropicbird nor my own routine excursions to collect data on the behaviour of this species and its strange association with hurricanes. Now, I am certain that the association is no accident, but it would have been foolhardy to mention my embryonic suspicions at the time.

A few months after Hector disappeared, I received notice that I had been accepted to a medical school in Houston. I graduated with honours in 2058, and entered a dual residency programme in neurosurgery and oncology. My professional expertise, and the work that consumes my waking hours, is the resection of brain

tumours. Don't ask me to speculate about reasons for the increased incidence of nervous-system cancers in Gulf Coast communities. I leave it to the researchers to argue about elevated ultraviolet exposures, pollution from submerged refineries and oncogene vectors released from inundated laboratories. My job, technical rather than scientific, is to remove the tumours surgically, and to try to leave the patient with memories and cognitive functions intact.

Which brings this story, ironically, to my current situation: greying hair partially shaved, scalp scrubbed and prepped with anaesthetics, sedatives flowing through my veins. It occurred to me that if memories can be localized and preserved during surgery, they can also be destroyed. Quite recently, I managed to convince two of my colleagues to attempt to obliterate a few of my memories, using a probe that I designed. Memories are the subunits of dreams: they remain locked, like political prisoners of a repressive cortical regime, in the hippocampus by day, released only when the rational brain lets down its guard during sleep. I no longer desire to understand the evolutionary origins of tropicbirds, or to discover why they suddenly appear in the Gulf, dozens of them, in advance of the frequent hurricanes. The longevity, unique DNA sequences and unfathomably odd neuroanatomy of these seabirds no longer interest me.

I just want the dreams to stop. ■
KV is a cancer-biology researcher, anatomical sciences instructor, artist and native of the Texas Gulf Coast. She very rarely writes fiction.

Join the discussion of Futures in Nature at go.nature.com/QMAM2

JACEY

Silk production from tarantula feet questioned

Arising from: S. N. Gorb *et al.* *Nature* **443**, 407 (2006)

As with all spiders, tarantulas spin silk from specialized structures in the abdomen called spinnerets, which are key features unique to the group. Recently Gorb *et al.*¹ reported that the zebra tarantula *Aphonopelma seemanni* also secretes silk from its feet, which might improve its ability to climb on vertical surfaces. Here we show that when the spinnerets are experimentally sealed, the zebra tarantula cannot secrete silk or similar threads, disagreeing with previous reports by Gorb *et al.*¹. Additional evidence also disagrees with leg secretion of silk.

We placed four specimens of *A. seemanni* each in separate glass containers with the floor and two vertical walls covered with 20 microscope slides. Tarantulas remained in the container for 14 h, after which the slides were carefully examined. Seventy-two hours later, we sealed the spinnerets of the same individuals with paraffin and repeated the experiments. In the first trials with free spinnerets most slides (24 out of 40 horizontal slides and 16 out of 40 vertical slides) showed silk threads together with dislodged abdominal urticating hairs. None of the slides in the trials with sealed spinnerets showed silk threads; we found only urticating hairs. In both series of trials we observed the spiders walking on vertical surfaces. In addition, we made transverse cuts of *A. seemanni* tarsi and no structures interpretable as silk glands or silk conduits were observed.

Tarantulas entangle silk threads from the spinnerets with their tarsi (Fig. 1). They often use hind legs to entangle silk, but we also observed



Figure 1 | The zebra tarantula *Aphonopelma seemanni* entangling a silk thread released from the spinneret with its hind leg, while starting to climb a vertical glass wall. Arrows indicate the silk thread.

A. seemanni entangling with their other legs. This behaviour might explain the presence of the thread footprints photographed by Gorb *et al.*¹. Tarantula silk released from spinnerets consists of multiple, roughly parallel silk threads which are better observed when magnified. This pattern could be maintained once the silk line is adhered to the substrate, when combed or stepped by tarsal scopulae. Tarsal structures interpreted as spigots by Gorb *et al.*¹ are very different from the diverse kinds of spigot reported for spiders^{2–4}. We found that the structures reported by Gorb *et al.*¹ are very similar in morphology and size to fragments of tarsal thermosensory setae reported for other tarantulas⁵. In insects, a common origin or convergence was proposed for silk spigots and sensory setae^{6,7}. The presence of conspicuous, dense scopulae with spatulated setae on the ventral surface of tarsi and metatarsi of all tarantulas seems incompatible with the presence of spinning spigots in a lower layer, because scopula setae might interfere with silk release. Furthermore, it has been explained that in *A. seemanni* the friction with scopula setae is enough to guarantee spider climbing⁸.

During our observations, individuals of *A. seemanni* brushed spinnerets with their hind-leg tarsi while walking and releasing silk threads. The presence of abdominal urticating hairs with silk is indicative of this brushing. We did not find evidence of silk-like secretion through leg structures in tarantulas.

Fernando Pérez-Miles¹, Alejandra Panzera¹, David Ortiz-Villatoro¹ & Cintya Perdomo¹

¹Sección Entomología, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay.

e-mail: myga@fcien.edu.uy

Received 29 December 2008; accepted 7 August 2009.

1. Gorb, S. N. *et al.* Silk-like secretion from tarantula feet. *Nature* **443**, 407 (2006).
2. Bond, J. E. Setae-spigot homology and silk production in first instar *Antrodiaetus unicolor* spiderlings (Araneae: Antrodiaetidae). *J. Arachnol.* **22**, 19–22 (1994).
3. Coddington, J. A. Spinneret silk spigot morphology: evidence for the monophyly of orbweaving spiders, Cyrtophorinae (Araneae), and the group Theridiidae plus Nesticidae. *J. Arachnol.* **17**, 71–95 (1989).
4. Griswold, C. E., Ramírez, M. E., Coddington, J. A. & Platnick, N. I. Atlas of phylogenetic data for entelegyne spiders. *Proc. Calif. Acad. Sci.* **56**, 1–324 (2005).
5. Raven, R. J. A new tarantula species from northern Australia (Araneae, Theraphosidae). *Zootaxa* **1004**, 15–28 (2005).
6. Ross, E. S. In *The Insects of Australia* 405–409 (Cambera & Melbourne Univ. Press, 1991).
7. Merritt, D. J. The organole concept of insect sense organs: Sensory transduction and organole evolution. *Adv. Insect. Physiol.* **33**, 192–241 (2007).
8. Niederegger, S. & Gorb, S. N. Friction and adhesion in the tarsal and metatarsal scopulae of spiders. *J. Comp. Physiol. A* **192**, 1223–1232 (2006).

doi:10.1038/nature08404

Gorb *et al.* reply

Replying to: F. Pérez-Miles, A. Panzera, D. Ortiz-Villatoro & C. Perdomo *Nature* **461**, doi:10.1038/nature08404 (2009)

The data presented by Pérez-Miles *et al.*¹ call into question the production of tarantula tarsal silk², but there remains substantive evidence that would seem to contradict their assertion that tarsal silk is solely due to abdominal silk entangled on to hind legs.

Tarsal silk traces appear consistently from all four pairs of legs, not just the hind pairs that might regularly contact the spinnerets. As can be seen in Fig. 2a of Gorb *et al.*², tarsal silk is laid down in parallel

tracks, which seems incompatible with it being a secondary deposition of silk combed on to the tarsi. Furthermore, scanning electron microscopy examination of tarsal silk reveals a distinct broad area at the beginning of the fibre that we interpret as an initial fluid (see Fig. 1). Presumably any fluid phase adhesive in abdominal silk would be lost in the transference to the tarsi. Another issue raised by Pérez-Miles *et al.*¹ is that adhesion of the spider to a clean dry surface could

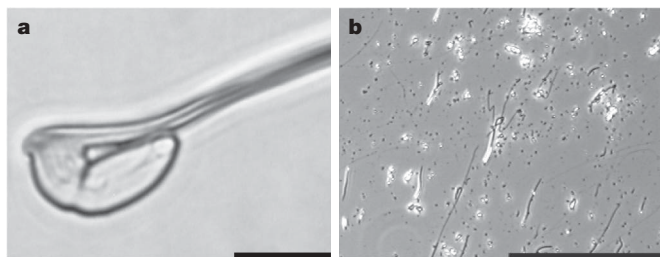


Figure 1 | Leg traces of *Aphonopelma seemanni* (light microscopy).
a, Starting point of a filamentous trace on coverslip. Scale bar, 1 μm .
b, Filamentous traces originating from such starting points. Scale bar, 100 μm .

be accomplished entirely with the tarsal setae. The tarsal claws would certainly also suspend the animal from very rough surfaces. The possibility of tarsal silk as a third adhesive mechanism implies that that the biology of tarantulas is such that there is an advantage to redundant systems.

Pérez-Miles *et al.*¹ also propose that the ‘spinneret’ structures we found are sensory structures, which is intriguing because similarity between the two structures may indicate a common developmental origin. Indeed, it has been suggested that the convergently evolved tarsal silks in web-spinner insects and dance flies arise from derived

sensory cells³. We remain satisfied that the most plausible explanation for our observations is the existence of tarsal-silk-producing structures scattered within the setae on tarantula tarsi.

Stanislav N. Gorb¹, Senta Niederegger², Cheryl Y. Hayashi³, Adam P. Summers⁴, Walter Vötsch⁵ & Paul Walther⁶

¹Zoological Institute of the University of Kiel, Am Botanischen Garten 1–9, D-24098 Kiel, Germany.

sgorb@zoologie.uni-kiel.de

²Institut für Rechtsmedizin, Universitätsklinikum Jena, Jena, Germany.

Senta.Niederegger@med.uni-jena.de

³University of California, Riverside, California 92521, USA.

⁴Friday Harbor Labs, University of Washington, Washington 98250, USA.

⁵Max Planck Institute for Developmental Biology, 71076 Tübingen, Germany.

⁶Electron Microscopy Department, University of Ulm, 89069 Ulm, Germany.

1. Pérez-Miles, F., Panzera, A., Ortiz-Villatoro, D. & Perdomo, C. Silk production from tarantula feet questioned. *Nature* **461**, 10.1038/nature08404 (2009).
2. Gorb, S. N. *et al.* Silk-like secretion from tarantula feet. *Nature* **443**, 407 (2006).
3. Merritt, D. J. The organule concept of insect sense organs: Sensory transduction and organule evolution. *Adv. Insect Physiol.* **33**, 192–241 (2007).

doi:10.1038/nature08405